

Uwe Kloos • Natividad Martínez • Gabriela Tullius

Tagungsband

Informatics Inside connect(IT);

20

Informatics
Inside



Hochschule Reutlingen
Reutlingen University

Tagungsband

Informatics Inside
connect(IT);

Uwe Kloos • Natividad Martínez • Gabriela Tullius

Tagungsband **Informatics Inside connect(IT);**

Informatik-Konferenz der Hochschule Reutlingen
20. Mai 2020

Herausgeber

Prof. Dr. Uwe Kloos	Prof. Dr. Natividad Martínez	Prof. Dr. Gabriela Tullius
Fakultät Informatik	Fakultät Informatik	Fakultät Informatik
Hochschule Reutlingen	Hochschule Reutlingen	Hochschule Reutlingen
Reutlingen, Deutschland	Reutlingen, Deutschland	Reutlingen, Deutschland



Hochschule Reutlingen
Reutlingen University

ISBN: 978-3-00-065431-2

Hochschule Reutlingen
Fakultät Informatik
Human-Centered Computing
Alteburgstraße 150
72762 Reutlingen

© 2020 Hochschule Reutlingen

Autoren: Jessica Konecki, Katharina Lohr, Adrian Schmidt, Antonia Schöning,
Sven Vennebusch und Dominik Walczak

Gestaltung, Illustration: Jessica Konecki

Lektorat: Jessica Konecki, Katharina Lohr, Adrian Schmidt, Antonia Schöning,
Sven Vennebusch und Dominik Walczak

Weitere Mitwirkende: Claudiu Bräuer, Yoan Dimitrov, Laura Discher, Alexander
Kahabka, Nicolai Kraus, Markus Rein, Wafa Sadri und Marc Viertel

infoinside@reutlingen-university.de
www.infoinside.reutlingen-university.de

Das Werk, einschließlich seiner Teile, ist urheberrechtlich geschützt. Jede Verwertung ist ohne Zustimmung des Verlages und des Autors unzulässig. Dies gilt insbesondere für die elektronische oder sonstige Vervielfältigung, Übersetzung, Verbreitung und öffentliche Zugänglichmachung.

Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie;
detaillierte bibliografische Daten sind im Internet über <http://dnb.d-nb.de> abrufbar.

Vorwort

Liebe Leserinnen und Leser,

zur diesjährigen Wissenschaftlichen-Vertiefungs-Konferenz (WVK) darf ich Sie herzlich begrüßen! Regelmäßig organisieren Studierende der Reutlinger Informatik die Konferenz mit äußerst hohem Engagement, um Ihnen eine Auswahl origineller wissenschaftlicher Arbeiten vorzustellen. Dabei stellen sich forschende Studierende im Rahmen ihres Studiums der Herausforderung, ihre Ergebnisse, die sie beispielsweise durch eine gewählte wissenschaftliche Vertiefung im Rahmen ihres Bachelor- oder Masterstudiums erhielten, anschaulich darzustellen.

Die Teilnehmer stellen die Ergebnisse ihrer schriftlichen Arbeit diesmal auf Grund der aktuellen Situation digital vor. Es ist zu wünschen, dass hierdurch sogar noch mehr Interessierte erreicht werden. Dass dies eine wertvolle Erfahrung für alle Beteiligten ist, steht außer Frage. Die WVK bietet ein Forum für Studierende während ihres Studiums, sowohl Ergebnisse ihrer Arbeit professionell einem interessierten Publikum zugänglich zu machen, als auch die Arbeiten anderer kritisch zu beurteilen und im Rahmen einer Konferenzorganisation zusammenzustellen. Dadurch erwerben die Studierenden weitere wichtige Qualifikationen, die über den Erwerb fachlicher Kenntnisse hinausgehen.

Anhaltend erlebt die Künstliche Intelligenz (KI) eine Renaissance in vielen Branchen. Der Trend, komplexe Zusammenhänge in Daten zu erfassen und zu nutzen, hält an. Hierbei ist jedoch der Grundgedanke des Maschinellen Lernens basierend auf empirischen Daten nicht neu. Es bleibt nach wie vor die Herausforderung, erst ein oft auch interdisziplinäres Verständnis von komplexen Zusammenhängen für verschiedenste Anwendungs-Domänen zu gewinnen, um zum Beispiel KI sinnvoll zum Einsatz zu bringen. Als Besucher der Konferenz erwarten Sie Beiträge aus den unterschiedlichsten Bereichen. Hierzu gehören zum Beispiel Müdigkeitserkennungssysteme im Automobil, ein Tastsinn auch für Roboter, aber auch neue Ansätze zur Erzeugung und Nutzung von Virtuellen Realitäten für die Erprobung des autonomen Fahrens bis hin zur Simulation von Außenboardeinsätzen in der Raumfahrt.

Ich wünsche allen Beteiligten und Besuchern eine interessante und inspirierende Konferenz.

Prof. Dr. Cristóbal Curio

Lehr- und Forschungsbereich Kognitive
Systeme, Vizedekan für Forschung an
der Fakultät Informatik

Reutlingen, Mai 2020

Inhaltsverzeichnis

Teil I Medizin

Computergestützte Lösungen für Herausforderungen in der Kryochirurgie	1
Nicolai Kraus	
Haptisches Feedback in der medizinischen Robotik	10
Adrian Schmidt	
Disease Detection and Health Monitoring Through Breath Analysis	21
Marc Viertel	
Benutzerstudie an einer Software von Kontrastmittelinjektoren für CT & MRT	29
Jessica Konecki	
Regulatorische Anforderungen an eine Qualitätsmanagement-Software für Medizinprodukte	41
Laura Discher	
Analyse des Requirements Engineering eines KIS-Herstellers	53
Katharina Lohr	
Traceability für Usability Engineering bei medizintechnischen Produkten	65
Antonia Schöning	

Teil II Neuronale Netze

Combining word embeddings and convolutional neural networks to detect duplicated questions	74
Yoan Dimitrov	
Visualisierung von neuronalen Netzen zur Unterstützung beim Training	83
Claudiu Bräuer	
Testumgebungen für Augmented Reality zur Anzeige autonomer Fahrfunktionen	95
Markus Rein	

Entwicklung eines Prototypen zur Müdigkeitserkennung im Fahr-simulator.....	106
Sven Vennebusch	

Teil III Weitere Themen

User Interaction and Motion Dynamics Simulations for Microgravity – Development of a Concept and Prototype for Astronaut Training with Virtual Reality	118
Wafa Sadri	

Vergleichende Analyse des YouTube-Auftritts von privat- und öffentlich-rechtlichen Sendegruppen	130
Alexander Kahabka	

Computergestützte Lösungen für Herausforderungen in der Kryochirurgie

Nicolai Kraus

Hochschule Reutlingen

Nicolai_Martin.Kraus@Student.Reutlingen-University.de

Abstract

In der Kryochirurgie wird Kälte verwendet, um tumoröses Gewebe abzutöten. Dazu werden Kryosonden in den Tumor gestochen und stark abgekühlt. Hierbei gibt es verschiedene Herausforderungen, welchen computergestützt begegnet werden kann. Diese Arbeit gibt die Ergebnisse einer Literaturrecherche zu den Herausforderungen wieder. Die vorgestellten Arbeiten beschäftigten sich mit der Simulation des im Tumor entstehenden Eisballs, dem korrekten Positionieren der Kryosonden im Tumor, dem Überwachen des Eingriffs sowie dem Entwickeln von Simulationen für Trainingszwecke. Dabei zeigt sich, dass der Einsatz von computergestützten Lösungen die Kryochirurgie für Operateur und Patient verbessern kann.

Schlüsselwörter

Kryochirurgie, Kryoablation, computergestützt, Simulation, Training

CR-Kategorien

• Human-centered computing ~ Visualization~ Visualization systems and tools

1 Einleitung

Kälte (Kryo) lässt sich in der Medizin vielseitig einsetzen und wird bereits seit Jahrzehnten angewendet [1, S. 9]. Ein Anwendungsbereich ist dabei die Kryochirurgie, welche u.a. gut- oder bösartige Tumore bekämpft. Dabei wird versucht den Tumor durch starke Einwirkung von Kälte zu zerstören. Mit der steigenden Anzahl an Tumorerkrankungen, was u.a. auf die älter werdende Bevölkerung zurückzuführen ist [5], lässt sich künftig eine steigende Anwendung der Kryochirurgie erwarten. Die sich stetig weiterentwickelnde Computertechnik lässt sich sinnvoll im Bereich der Kryochirurgie einsetzen und ermöglicht viel Optimierungspotential. Diese Arbeit soll Grundlagen der Kryochirurgie, sowie computergestützte Anwendungen in der Kryochirurgie vorstellen.

2 Kryochirurgie

Dieser Abschnitt nennt Grundlagen der Kryochirurgie. Dabei werden Anwendungsbereiche, Vor- und Nachteile, Ablauf eines Eingriffs, sowie die Funktionsweise der kryochirurgischen Komponenten aufgezeigt.

141

Betreuer Hochschule: Prof. Dr.-Ing. Cristóbal Curio
Hochschule Reutlingen
Cristobal.Curio@Reutlingen-University.de

Wissenschaftliche Vertiefungskonferenz
20. Mai 2020, Hochschule Reutlingen
Copyright 2020 Nicolai Kraus

2.1 Anwendungsbereiche

Die Kryochirurgie hat mehrere Anwendungsbereiche, wobei der Schwerpunkt dieser Arbeit auf der Zerstörung von Tumoren durch Einbringen und Abkühlen von Sonden in einem Tumor liegt. Ein weiterer Anwendungsbereich ist die Kryobiopsie, wobei eine Sonde in die Lunge geschoben an eine gewünschte Position gebracht und stark abgekühlt wird. An der Sondenspitze friert Lungengewebe fest, wird herausgezogen und kann anschließend untersucht werden [2]. Ein weiterer Anwendungsbereich ist in der Kardiologie zu finden. Dabei wird ein Kryoballon zur Bekämpfung von Vorhofflimmern im Herzen verwendet [10].

2.2 Vergleich zur konventionellen Chirurgie

Im Vergleich zu einem konventionellen chirurgischen Einsatz hat die Kryochirurgie Vor- aber auch Nachteile. So sind die Eingriffe meist minimalinvasiv und können auch ambulant durchgeführt werden, was wiederum kaum Narben verursacht. Ebenso werden bei dieser Technik kaum Gerüststrukturen im Körper zerstört, im Vergleich zu einem chirurgischen Eingriff, bei dem Gewebe entfernt wird. Des Weiteren kann die Kryochirurgie auch an schwer erreichbaren Körperregionen eingesetzt werden. Beim Faktor Kosten ist die Kryochirurgie kostengünstiger als die konventionelle Chirurgie [1, S. 13]. Dadurch, dass bei der Kryochirurgie das zerstörte Gewebe im Körper verbleibt, ist es nicht möglich eine Gewebeprobe zu entnehmen und zu prüfen, ob die Abkühlung ausreichend war, um das betroffene Gewebe vollständig zu zerstören [1, S. 22-23].

2.3 Ablauf

Ein kryochirurgischer Eingriff zum Bekämpfen eines Tumors benötigt eine präzise Therapieplanung. Ein entscheidender Punkt ist dabei, wie lange Kälte an den Tumor abgegeben werden muss, um diesen vollständig zu zerstören. Ebenso ist zu beachten, dass ein Bereich, der vereist worden ist, nicht automatisch bedeutet, dass dieses Gewebe auch zerstört wurde [1, S. 22-23].

Der Ablauf eines kryochirurgischen Eingriffs ist grundlegend wie folgt aufgebaut: Der erste Schritt ist die Planung. Hier werden bildgebende Verfahren wie die Magnetresonanztomographie (MRT) oder der Ultraschall verwendet, um den Ort des Eingriffs beurteilen zu können. Anschließend werden die Kryosonden im Tumor positioniert, mit anschließender Kontrolle der Position mithilfe der bildgebenden Verfahren. Danach kann die Kälteanwendung beginnen. Dabei entsteht der sogenannte Eisball an der Sondenspitze. Der Eisball hat in der Kryochirurgie eine essenzielle Rolle, weshalb sich mit der Form und der Entstehung einige der nachfolgenden Arbeiten intensiv beschäftigen. Die Kälteanwendung umfasst dabei um die 15 Minuten der Kühlung und eine 10-minütige Auftauphase. Dieser Vorgang wird zwei- bis dreimal wiederholt [7].

Während des Eingriffs ist zum Abtöten von Tumorzellen eine Temperatur von -40°C nötig. Ebenso ist ein langsamer Auftauprozess notwendig, weil dieser ebenfalls zur Zellzerstörung beiträgt [1, S. 10].

2.4 Funktionsweise

Ein System für eine kryochirurgische Anwendung lässt sich durch wenige Komponenten realisieren. Dazu gehört eine Gasflasche, welche das Kältemittel enthält, eine oder mehrere Sonden, in die das Kältemittel der Gasflasche geleitet wird, sowie eine Möglichkeit, das Kältemittel aus der Gasflasche zu steuern, z. B. über einen Fußschalter [2]. Die Art der verwendeten Sonde, welche die Kälte später in den Körper abgeben soll, ist stark vom Anwendungsfall abhängig. So können in der Kryobiopsie flexible Sonden angewendet werden [2], starre Sonden dagegen zum Einbringen in einen Tumor. Die Durchmesser der Sonden befinden sich dabei in Bereichen um die 2 mm. Durch diese Sonde (Abbildung 1) wird ein Gas wie z. B. Distickstoffoxid (N_2O), Kohlenstoffdioxid (CO_2) oder Argon (Ar) geleitet. Der Joule-Thomson-Effekt, der entsteht, wenn es bei den Gasen zu einer Druckänderung kommt (Gasentspannung), führt zu einer starken Abkühlung an der Sondenspitze [1, S.16-17] [3].

Die Stärke der Abkühlung ist abhängig vom verwendeten Gas. So können mit CO_2 Temperaturen um die -78°C erreicht werden, mit N_2O -89°C und mit Ar -160°C [1, S.16].

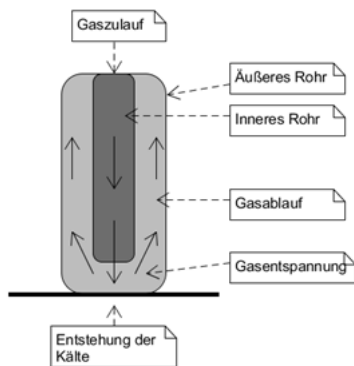


Abbildung 1: Aufbau einer Kryosonde,
Quelle: In Anlehnung an [1, S.17]

3 Herausforderungen

Kryochirurgische Eingriffe beinhalten verschiedene Herausforderungen, wie z. B. das korrekte Positionieren der Kryosonde oder die richtige Dauer der Kälteabgabe. Verschiedene Arbeiten beschäftigen sich damit, wie diese Schwierigkeiten computergestützt reduziert werden können.

3.1 Eissimulation unter Berücksichtigung von Blutgefäßen

Bereits 1999 erwähnte die Arbeit von Buschmann et al. [1, S. 23] die Schwierigkeit der Berechnung der entstandenen Temperaturfelder bei der Kryoanwendung. Verschiedene Veröffentlichungen haben sich dabei mit dem Problem numerisch und analytisch auseinandergesetzt. Dabei stellte sich heraus, dass die Hauptschwierigkeit das sehr individuelle Gewebe ist, welches eine genaue Berechnung stark erschwert.

Dennoch gibt es heute Arbeiten [4], die sich mit diesem Problem weiterhin auseinandersetzen. Dabei werden meistens computergestützte Simulationen des Eisballs für den

kryochirurgischen Einsatz erzeugt. Eine der größten Herausforderungen dabei ist das betroffene und das umliegende Gewebe, welche die Entstehung des Eisballs beeinflussen. Mit dem Ziel den Tumor vollständig zerstören zu können ist es notwendig, alle Einflussfaktoren zu berücksichtigen, damit der Eisball den vollständigen Tumor abdecken kann. Relevant ist zudem, ob für die Kryoablation des Tumors eine oder mehrere Sonden verwendet werden, was das Resultat ebenfalls beeinflusst [4].

Die Arbeit von Golkar et al. [4] beschäftigt sich mit der 3D Simulation eines Eisballs bei mehreren Kryosonden unter der Berücksichtigung eines Gefäßes, welches sich in der Gefrierzone befindet. Für die Berechnung stellten Golkar et al. [4] eine partielle Differentialgleichung auf, um die Wärmeausbreitung für einen bestimmten Zeitbereich in einem 3D Koordinatensystem zu beschreiben. Die Gleichung passten sie entsprechend an, sodass durch das Volumen ein Blutgefäß gelegt werden kann. Alle Voxel, die von dem Gefäß berührt wurden, behielten während der Simulation dieselbe Temperatur. Golkar et al. [4] legten für ihre Simulation diverse Parameter, wie z. B. die Auflösung in ihrem Volumenmodell auf 1 mm fest, aber auch, dass die Kälteabgaben der Kryosonden konstant bleiben. Die Implementierung erfolgte mithilfe des CUDA Toolkit von NVIDIA mit dem Fokus auf eine schnelle Berechnung mithilfe des Grafikprozessors. Durch zwei unterschiedliche Methoden prüften Golkar et al. [4] ihre entwickelte Simulation. Mit dem ersten Ansatz prüften sie dabei die grundsätzliche Korrektheit ihrer Simulation. Dabei simulierten sie das Wachstum des Eisballs und verglichen das Resultat mit der Arbeit von Shah et al. [9], die das Wachstum in einem Ultraschallgel durchführten. Ein Vergleich der Eisbälle zwischen den beiden Arbeiten zeigte in der Simulation eine Abweichung von nur 5,8 %. Anhand eines echten kryochirurgischen Eingriffs an der Niere, der mit einem MRT mitverfolgt wurde, wurde die Simulation ein zweites Mal geprüft. Das Entscheidende hierbei war, dass der Eingriff in der Nähe eines Gefäßes verlief. Um den Eingriff

auf die Simulation übertragen zu können, wurden die MRT-Bilder, die den Verlauf des Eingriffs aufzeigten, zueinander registriert und anschließend segmentiert. Die Positionen der Kryosonden in den MRT-Bildern konnten entsprechend auf die Simulation übertragen werden. Ähnlich wie im MRT zeigte auch die Simulation deutliche Veränderungen in Bereichen, in denen die größeren Blutgefäße lagen. Unter anderem beurteilen Golkar et al. [4] mithilfe des Dice Similarity Coefficient (DSC) die Unterschiede, die sich nur auf 0,82 beliefen. Je näher die Zahl des DSC dabei an 1 ist, desto ähnlicher sind die Ergebnisse. Abbildung 2 zeigt die Simulation der Arbeit. Die Wolke links im Bild zeigt dabei die Gefäße, welche sich in der Nähe des simulierten und von den MRT-Bildern segmentierten Eisballs befinden. Weißlich und teilweise überdeckt dargestellt ist der segmentierte Eisball aus den MRT-Bildern.

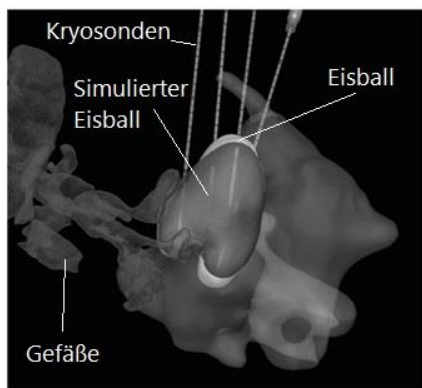


Abbildung 2: Simulation des Eisballs an den Sonden, Überarbeitete Darstellung von Quelle: [4, S. 236]

Eine weitere hier erwähnenswerte Arbeit ist von Essert et al. [12], welche sich mit der Simulation der entstehenden Kälte im hinteren Bereich der Kryosonden beschäftigen. Dabei kann die Abkühlung der Kryosonde im hinteren Bereich zu einer Schädigung von tumorfreiem Gewebe führen.

3.2 Positionieren der Kryosonden

Nachdem die Therapieplanung abgeschlossen ist, kann mit dem Eingriff begonnen werden. Hierbei müssen die Kryosonden an der betroffenen Stelle so genau wie möglich positioniert werden, was durch verschiedene bildgebende Verfahren unterstützt werden kann. Im Falle der Verwendung eines MRTs kann dieser Prozess sehr langwierig sein, da nach jeder Kryosondenpositionierung ein MRT durchgeführt werden muss, um zu prüfen, ob die Kryosonde korrekt gesetzt wurde. Der Patient muss dabei jedes Mal aus der MRT-Röhre geholt und wieder hineingeschoben werden [13].

Der Schwierigkeit einer genauen Positionierung der Kryosonden nimmt sich die Arbeit von Gao et al. [13] an. Dabei wurde ein System entwickelt, welches in Echtzeit die Position der Kryosonden optisch und elektromagnetisch ermittelt, und somit die Positionierung der Kryosonden unterstützen kann. Um dies zu erreichen wurden an den Kryosonden Marker (Fiducials) für die optische Erkennung sowie ein Sensor für die elektromagnetische Erkennung angebracht. Der elektromagnetische Sensor war dabei so gewählt, dass er im Magnetfeld des MRTs arbeiten konnte. Dadurch verwendeten das MRT als auch der elektromagnetische Sensor dasselbe Koordinatensystem und eine spätere Registrierung zueinander entfiel. Der Grund für die Kombination von optischer sowie elektromagnetischer Erkennung war, dass bei einer rein optischen Erkennung die Sicht auf die Marker zeitweise durch den Arzt beim Arbeiten verdeckt sein kann und somit eine Lokalisierung nicht mehr gegeben ist. Gao et al. [13] verwendeten für ihren Versuch ein Wirbelsäulenattrappe in Gelatine. Daran wurden Marker angebracht und anschließend MRT-Aufnahmen erzeugt. Damit konnten die verschiedenen Koordinatensysteme der unterschiedlichen Komponenten wie optische und elektromagnetische Marker sowie die MRT-Aufnahmen zueinander transformiert werden. Die Visualisierung erfolgte hierbei mithilfe der Software 3D Slicer. Um die Genauigkeit ihres Systems zur Kryosondenpositionierung

zu prüfen, positionierten sie mehrere Kryosonden in der Attrappe und prüften diese diesmal mithilfe von Computertomographie (CT)-Aufnahmen, aufgrund der höheren Genauigkeit im Vergleich zu einem MRT. Als Ergebnis zeigte sich, dass die Positionierungsfehler der rein optischen Erkennung bei 4,23 mm, der elektromagnetischen bei 2,21 mm und der Kombination aus beiden bei 1,87 mm lag. Die vermutet höhere Genauigkeit der optischen Erkennung aufgrund der Spezifikationen war nicht zutreffend. Dies ließ sich damit erklären, dass die Kryosonden nicht komplett starr waren und der optische Marker weiter hinten als der elektromagnetische Marker an der Kryosonde befestigt war. Somit war der Spielraum bei einem leichten Verbiegen, während der Anwendung der Kryosonde für den optischen Marker größer als für den elektromagnetischen.

Die Arbeit von Gao et al. [13] zeigt damit ein System zum genauen Positionieren von Kryosonden in Echtzeit. Durch die Verwendung der elektromagnetischen und optischen Verfolgung konnte ein genaues und permanentes Positionieren ermöglicht werden, trotz der Verdeckung der Sicht für den optischen Sensor.

3.3 Fortschrittskontrolle während eines Eingriffs

Nicht nur beim Positionieren der Kryosonden, sondern auch während des kryochirurgischen Eingriffes ist es notwendig, den Fortschritt des sich bildenden Eisballs mitzuverfolgen, wobei sich das MRT als bildgebendes Verfahren anbietet. Da der Arzt während solch eines Eingriffs u.a. damit beschäftigt sein kann die Unterkühlungen an anderen Körperstellen zu verhindern, bietet es sich an, diese Überwachung automatisiert erfolgen zu lassen [7]. Mit dieser Herausforderung beschäftigt sich die Arbeit von Liu et al. [7], wobei das Ziel eine automatische Segmentierung des Eisballs während eines Eingriffs mithilfe eines MRTs ist. Da sowohl der Eisball als auch das umliegende Gewebe im MRT ähnliche Helligkeitswerte aufweisen, ist diese Aufgabe mit alleiniger Verwendung dieses Parameters nicht möglich [7]. Die von Liu et al. [7] verwendeten MRT-Daten

zum Entwickeln und Testen ihrer Segmentierung beinhalteten 13 Eingriffe in der Niere zum Bekämpfen eines Tumors. Ein derartiger Eingriff besteht aus dem Positionieren der Sonden im Tumor, sowie mehrfachen Gefrierzyklen zum Abtöten des betroffenen Gewebes. Die Dauer eines Gefrierzyklus beträgt dabei 15 Minuten, welcher alle 3 Minuten mithilfe eines MRTs geprüft wird. Für die Entwicklung des Algorithmus war es notwendig zuerst herauszufinden, wie sich ein Eisball in Realität unter Optimalbedingungen entwickelt. Um eine ungefähre Vorgabe der Eisballentwicklung nach den verschiedenen Zeiträumen zu haben, wurde die Eisballentstehung in einem Gelbad durchgeführt und dabei Aufnahmen mit einem CT gemacht. Daraus konnte ein Modell entwickelt und bestimmt werden, das den Wachstumsprozess zu beliebigen Zeiten beschreibt. Der Ablauf des Segmentierungsalgorithmus wird in Abbildung 3 gezeigt. Ausgangspunkt ist dabei der MRT-Scan (Links in Abbildung 3), der ausgeführt wurde, nachdem alle Kryosonden im Tumor positioniert wurden. Die Erkennung der Kryosonden im MRT-Bild erfolgt dabei automatisiert und wird in einer vorausgehenden Arbeit von Liu et al. [11] beschrieben. Durch die Positionen der einzelnen Kryosonden kann anschließend mithilfe des vorher entwickelten Modells die Form des Eisballs (Eisballmodell) zu beliebigen Zeitpunkten berechnet werden. Auf der rechten Seite in Abbildung 3 befindet sich der erste Vorverarbeitungsschritt an den MRT-Bildern. Die MRT-Bilder wurde mithilfe eines Fuzzy-c-Means-Algorithmus [14] jedem Voxel die Kategorie Eisball oder umliegendes Gewebe zugeordnet [7].

Mit dem Wissen der Eisballform durch die positionierten Kryosonden, sowie der ersten Segmentierung mithilfe des Fuzzy-c-Means-Algorithmus im MRT-Bild wird anschließend mit dem Fast Marching Propagation-Algorithmus [15] die Form des Eisballs angepasst. Der Algorithmus arbeitet iterativ, dabei wächst und schrumpft der zu segmentierende Bereich, bis er die festgelegten Grenzwerte erreicht hat. Anschließend wird mithilfe des Shape Priori Modeling geprüft, wie wahrscheinlich es ist,

dass ein Voxel auf Grund seiner Position wirklich zum Eisball gehört oder nicht. Davon gefolgt wird mithilfe des GraphCut-Algorithmus das Bild in die Bereiche Objekt, Eisball, Hintergrund und das umliegende Gewebe unterteilt [7]. Die Qualität des entwickelten Segmentierungsalgorithmus prüften Liu et al. [7] indem sie ihn mit einer händischen Segmentierung abglichen. Mithilfe des DSC wurde geprüft, wie ähnlich die Ergebnisse waren. Nach einem Vergleich aller 13 Eingriffe konnte ein DSC-Wert von 0.92 erreicht werden.

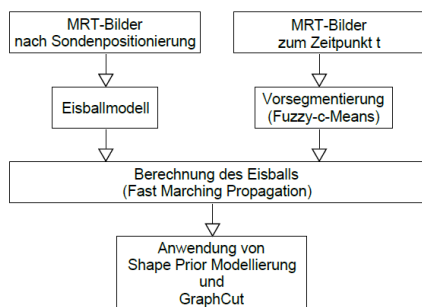


Abbildung 3: Ablauf der Segmentierung des Eisballs, Quelle: In Anlehnung an [7, S. 518]

3.4 Artefakterzeugung in Ultraschallbildern

Kryochirurgische Eingriffe sind komplex und benötigen Erfahrung. Eine Herausforderung dabei ist die korrekte Positionierung der Kryosonden, was im Falle eines Eingriffs in der Prostata mit einem Transrektalen Ultraschall überwacht wird. Mithilfe von Simulationen können diese Eingriffe trainiert werden [6]. Mit diesem Themenbereich beschäftigt sich die Arbeit von Keelan et al. [6]. Dabei geht es speziell um die Generierung von Artefakten in einem zuvor aufgenommenen Ultraschallbild. Diese Artefakte sind zum Trainieren eines realistischen kryochirurgischen Eingriffs notwendig. Die Artefakte entstehen dabei durch die eingebrachten Kryosonden, aber auch durch die entstandene gefrorene Region während des Abkühlprozesses [6]. Die zwei Fra-

gen bei dieser Arbeit waren, wie diese Artefakte realistisch und zudem auch in Echtzeit für eine Simulation eines kryochirurgischen Eingriffs erzeugt werden können. Dabei gibt es Ansätze wie interpolative Verfahren, welche jedoch keine komplexe realistische Darstellung mehrerer Reflektionen berücksichtigen. Ein anderer Ansatz wäre die Verwendung von Ray Tracing, wobei ein erzeugter Strahl nachverfolgt wird und entsprechend auf die Oberfläche reagiert, auf die er aufrifft. Hier ist jedoch das Problem, dass im Falle von Ultraschalldaten das Ray Tracing nicht gut anwendbar ist, da die Strahlen zu sehr durch die Artefakte des Gewebes im Ultraschallbild abgelenkt werden. Keelan et al. [6] entschieden sich deshalb dafür eine Kombination aus interpolativen Verfahren sowie dem Ray Tracing zu realisieren. Die Simulation sollte damit in der Lage sein, Artefakte der Kryosonden anzeigen zu können, sowie die gefrorenen Bereiche während der Anwendung darzustellen [6].

Die Artefakterzeugung im Ultraschallbild wird von Keelan et al. [6] in sechs Berechnungsschritte unterteilt. Dabei wird zuerst berechnet, wie sich die Kälte in dem betroffenen Bereich während des Eingriffs entwickelt. Mithilfe dieses Ergebnisses wird geprüft, wie dies die Ultraschallsignale beeinflusst. Im dritten und vierten Schritt wird berechnet, wie sich die Ultraschallsignale im Gewebe ausbreiten und abgelenkt werden, und wie diese an Energie verlieren, wenn sie die verschiedenen Gewebeschichten durchdringen. In den letzten beiden Schritten wird mithilfe des Energiewertes des zurückkommenden Strahls das Ultraschallbild und dessen Helligkeitswerte errechnet, und anschließend in die ursprünglichen Ultraschallaufnahmen übertragen. Ebenso entscheidend bei dieser Arbeit war die Möglichkeit einer Echtzeitdarstellung der Simulation, wobei sich die Anwendung des Ray Tracings als am rechenintensivsten herausstellte. Entsprechend wurden verschiedene Optimierungen in Richtung der Verwendung des Grafikprozessors gemacht. Algorithmisch wurden zudem berechnete Ultraschallstrahlen in der Simulation entfernt, welche aufgrund ihrer starken Abschwächung den Weg zurück zum

Sender (im echten Anwendungsfall die Ultraschallsonde) nicht erreichen können, womit die Berechnung um den Faktor zwei beschleunigt werden konnte [6].

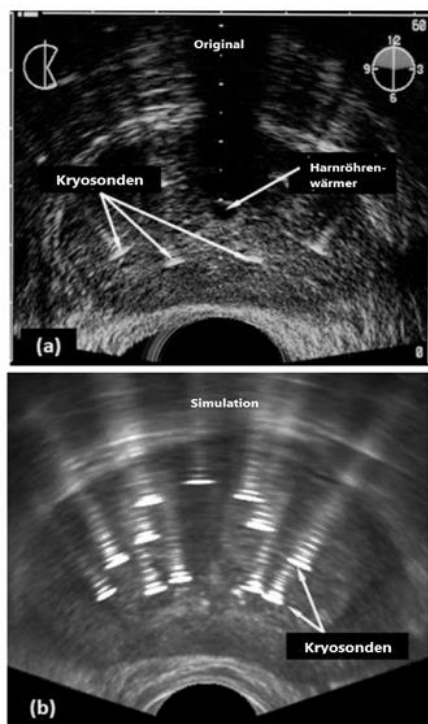


Abbildung 4: Artefakte durch Kryosonden in Ultraschallbild. Abbildung (a) zeigt Originalaufnahmen und (b) die Simulation, Quelle: [6, S. 11] (Beschriftungen wurden übersetzt)

Der in der Arbeit verwendete gewöhnliche Arbeitscomputer und der Algorithmus erreichten schlussendlich 100 Bilder pro Sekunde, womit eine Echtzeitdarstellung problemlos möglich war. Die in Abbildung 4 gezeigten transversalen Ultraschallaufnahmen zeigen einen Teil der Resultate der Arbeit von Keelan et al. [6]. Dabei zeigt die obere Abbildung 4 (a) originale Ultraschallaufnahmen eines kryochirurgischen Eingriffs und die dabei verwendeten Kryosonden, und darunter in Abbildung 4 (b) Ultraschallaufnahmen, bei denen die Artefakte

von dem Algorithmus einberechnet wurden. Die Unterschiede in der Bildqualität sind darauf zurückzuführen, dass das Ultraschallgerät, mit dem die obere Abbildung 4 (a) erzeugt wurde, eine niedrigere Auflösung hatte als die Aufnahme (b), in die die Artefakte für die Simulation integriert wurden [6].

Ebenfalls ermöglicht die Simulation die Artefakterzeugung in sagittaler Ansicht. Damit zeigten Keelan et al. [6] eine realistische und in Echtzeit berechenbare Simulation von Artefakten in Ultraschallbildern für Trainingssysteme im Bereich der Kryochirurgie.

4 Diskussion

Diese Arbeit hat verschiedene Veröffentlichungen im Bereich der Kryochirurgie vorgestellt. Dabei war der Fokus auf Herausforderungen, denen computergestützt begegnet werden kann, gelegt. Bei der Literaturrecherche zeigte sich anhand der Anzahl an Veröffentlichungen, dass sich viele Arbeiten bei diesem Themenbereich vor allem mit der Berechnung der Kälteausbreitung befassen. Einer dieser Ansätze wurde in dieser Arbeit näher in Kapitel 3.1 beschrieben.

Weitere bei der Literaturrecherche gefundene Arbeiten zeigten, dass die Herausforderungen sehr vielseitig sind und sich auch mit sehr spezifischen Teilbereichen der Kryochirurgie beschäftigen. Dazu gehört z. B. das automatische Erkennen der Kryosonden in MRT-Bildern [11] aber auch die Artefakterzeugung in Ultraschallbildern [6].

Die Herausforderungen denen computergestützt begegnet werden kann, lassen sich, bezogen auf diese Arbeit, in folgende Bereiche unterteilen:

- Trainieren von Eingriffen
- Vorbereiten von Eingriffen
- Überwachen von Eingriffen

Vor allem im Bereich des Trainings von kryochirurgischen Eingriffen gibt es mehrere Veröffentlichungen. Die hier vorgestellte Arbeit von Keelan et al. [6] fokussiert sich dabei auf

einen sehr speziellen Bereich wie der Artefakterzeugung in Ultraschallbildern, welche später in Trainingssimulatoren implementiert werden können. Andere, hier jedoch nicht vorgestellte Arbeiten wie von Rabin et al. [8] zeigten dabei ganze Tutorsysteme für kryochirurgische Eingriffe im Bereich der Prostata.

Eine ältere und dementsprechend auch in vielen Veröffentlichungen behandelte Disziplin in der Kryochirurgie ist das Berechnen oder Simulieren des Eisballs. Hierbei können durch leistungsstarke 3D Simulationen die Eisbälle unter Berücksichtigung von Blutgefäßen berechnet werden, wie die Arbeit von Golkar et al. [4] zeigt. Mithilfe dieser Simulationen lassen sich kryochirurgische Anwendungen gezielt im Voraus planen und an die Gegebenheiten des Patienten anpassen, womit bessere Behandlungsergebnisse für den Patient erreicht werden können.

Herausfordernd in der Kryochirurgie ist vor allem, dass der Eingriff erfolgt ohne direkt das Ziel, den Tumor, zu sehen, weshalb auch während des Eingriffs eine Überprüfung des Vorgangs durch den Arzt notwendig ist. Die Arbeit von Liu et al. [7] zeigt dabei, wie dies aufgrund von MRT-Bildern automatisiert erfolgen kann. Mit steigender Qualität der bildgebenden Verfahren lassen sich auch hier zukünftig immer bessere Ergebnisse erwarten, die dem Patienten zugutekommen.

Ebenfalls fällt bereits bei den Titeln und später in der Arbeit von Golkar et al. [4] sowie Keelan et al. [6] der Fokus ihrer Algorithmen auf eine schnelle Berechnung mithilfe des Grafikprozessors, auf. Dies lässt sich mit der Notwendigkeit einer Echtzeitdarstellung aber auch dem aufwendigen Berechnen von Volumenmodellen erklären.

Zukünftig lassen sich Kombinationen der hier aufgezeigten Arbeiten erwarten, wie z. B., dass nach dem Positionieren der Kryosonden [13], anschließend eine Simulation des Eisballs [4], durchgeführt wird.

5 Fazit

Die Kryochirurgie kann in Bereichen wie der Bekämpfung von Tumoren alternativ zur klassischen Chirurgie sinnvoll eingesetzt werden.

Dadurch, dass ein kryochirurgischer Eingriff minimalinvasiv erfolgt und eine freie Sicht auf den Tumor nicht immer möglich ist, können verschiedene computergestützte Ansätze die Anwendung und somit das Resultat verbessern. Dabei lassen sich durch Simulationen Gefäße einberechnen, die die betroffene Ablationszone beeinflussen, wodurch ein Eingriff besser vorausgeplant werden kann. Aber auch während des Eingriffs können Verfahren, wie die automatische Segmentierung des Eisballs in MRT-Bildern angewendet werden. Ebenso zeigte sich, dass das Trainieren eines kryochirurgischen Eingriffs eine hohe Bedeutung hat.

6 Literaturverzeichnis

- [1] W. Buschmann. Kryochirurgie von Tumoren in der Augenregion: Weiterentwicklungen der Kryochirurgie in Ophthalmologie und Dermatologie. Enke, Stuttgart, 1999.
- [2] S. Colella. Cryobiopsy: Physics. In V. Poletti, editor, Transbronchial cryobiopsy in diffuse parenchymal lung disease, pages 41–45. Springer International Publishing, Cham, 2019.
- [3] J. P. Erinjeri and T. W. I. Clark. Cryoablation: mechanism of action and devices. *Journal of vascular and interventional radiology: JVIR*, 21(8Suppl): S187–91, 2010.
- [4] E. Golkar, P. P. Rao, L. Joskowicz, A. Gangi, and C. Essert. Fast gpu computation of 3d isothermal volumes in the vicinity of major blood vessels for multiprobe cryoablation simulation. In A. F. Frangi, J. A. Schnabel, C. Davatzikos, C. Alberola-López, and G. Fichtinger, editors, *Medical Image Computing and Computer Assisted Intervention–MICCAI 2018*, volume 11073 of *Lecture Notes in Computer Science*, pages 230–237. Springer International Publishing, Cham, 2018.
- [5] W. Hiddemann, M. Feuring-Buske, L. Lindner, M. Krych, H. Huber, and C. R.

- Bartram. Was ist krebs? In W. Hidde-
mann and C. R. Bartram, editors, *Die
Onkologie*, pages 3–16. Springer Berlin
Heidelberg, Berlin, Heidelberg, 2010.
- [6] R. Keelan, K. Shimada, and Y. Rabin.
Gpu-based simulation of ultrasound im-
aging artifacts for cryosurgery training.
*Technology in cancer research & treat-
ment*, 16(1):5–14, 2017
- [7] X. Liu, K. Tuncali, W. M. Wells, and G.
P. Zientara. Automatic iceball segmenta-
tion with adapted shape priors for mri-
guided cryoablation. *Journal of magnetic
resonance imaging : JMRI*, 41(2):517–
524, 2015.
- [8] Y. Rabin, K. Shimada, P. Joshi, A. Seh-
rawat, R. Keelan, D. M. Wilfong, and J.
T. McCormick. A computerized tutor
prototype for prostate cryotherapy: Key
building blocks and system evaluation.
*Proceedings of SPIE—the International
Society for Optical Engineering*, 10066,
2017.
- [9] T. T. Shah, U. Arbel, S. Foss, A. Zach-
man, S. Rodney, H. U. Ahmed, and M.
Arya. Modeling cryotherapy ice ball di-
mensions and isotherms in a novel gel-
based model to determine optimal cryo-
needle configurations and settings for
potential use in clinical practice. *Uro-
logy*, 91:234–240, 2016.
- [10] F. Straube, U. Dorwarth, and E. Hoff-
mann. *Fortschritte der kryoballontech-
nik: Nutzen und risiken in der täglichen
anwendung. Herzschrittmachertherapie
& Elektrophysiologie*, 25(4):230–235,
2014
- [11] X. Liu, K. Tuncali, WM. Wells III, GP
Zientara. Automatic probe artifact detec-
tion in MRI-guided cryoablation. *Pro-
ceedings of SPIE Medical Imaging: Im-
age-Guided Procedures, Robotic Inter-
ventions, and Modeling*; Lake Buena
Vista, Florida, 2013.
- [12] C. Essert, P.P. Rao, A. Gangi, L. Jos-
kowicz (2019) 3D Modelling of the Re-
sidual Freezing for Renal Cryoablation
Simulation and Prediction. In: Shen D. et
al. (eds) *Medical Image Computing and
Computer Assisted Intervention – MIC-
CAI 2019. MICCAI 2019. Lecture Notes
in Computer Science*, vol 11768.
Springer, Cham
- [13] W. Gao, B. Jiang, D. F. Kacher, B.
Fetics, E. Nevo, T. C. Lee, & J.
Jayender, (2018). Real-time probe track-
ing using EM-optical sensor for MRI-
guided cryoablation. *The international
journal of medical robotics + computer
assisted surgery: MRCAS*, 14(1),
10.1002/rcs.1871.
- [14] D. Pham. (2001). Spatial Models for
Fuzzy Clustering. *Computer Vision and
Image Understanding*. 84. 285-297.
10.1006/cviu.2001.0951.
- [15] J. A. Sethian, *Level set methods and fast
marching methods: Evolving interfaces
in computational geometry, fluid me-
chanics, computer vision*, Cambridge
University Press, 1999

Haptisches Feedback in der medizinischen Robotik

Adrian Schmidt
Hochschule Reutlingen

Adrian.Schmidt@Student.Reutlingen-University.de

Abstract

Haptisches Feedback ist nach zahlreichen Studien ein wichtiger Bestandteil in der medizinischen Robotik. Die meisten Systeme befinden sich jedoch noch im Forschungsstatus und verfolgen unterschiedliche Ansätze. In der Teleoperation wird mit sensorlosen und Sensor-Systemen geforscht. Sensoren bieten, im Gegensatz zu den Encodern in sensorlosen Systemen, genaue Messungen, sind allerdings teuer in der Anschaffung, schwer zu desinfizieren und müssen in OP-Besteck integriert werden. In Hands-On Systemen fühlt der Operateur im Gegensatz zu Teleoperationssystemen direkt die auftretenden Kräfte bei der Benutzung. Der Roboter bietet in diesen Systemen nur die benötigte Stabilität und Genauigkeit, gesteuert werden sie direkt durch den Menschen. Dagegen werden in Teleoperationssystemen spezielle Controller eingesetzt. Hier hat sich der für den OP entwickelte sigma.7 durchgesetzt. Gegenüber der für die Allgemeinheit entwickelten Konkurrenz bietet er haptisches Feedback in allen nötigen Freiheitsgraden und eine entsprechende Kraftrückkopplung.

Betreuer Hochschule: Prof. Dr.-Ing. Oliver Burgert
Hochschule Reutlingen
Oliver.Burgert@Reutlingen-University.de

Wissenschaftliche Vertiefungskonferenz
20. Mai 2020, Hochschule Reutlingen
Copyright 2020 Adrian Schmidt

Schlüsselwörter

Haptisches Feedback, Robotik, Teleoperation, Sensorik, Aktorik

CR-Kategorien

- Computer systems organization
- ~Embedded and cyber-physical systems
- ~Robotics
- ~External interfaces for robotics
- Computer systems organization
- ~Embedded and cyber-physical systems
- ~Robotics
- ~Robotic control
- Computer systems organization
- ~Embedded and cyber-physical systems
- ~Sensors and actuators

1 Einleitung

1.1 Motivation und Ziele

Haptisches Feedback in der medizinischen Robotik bietet einige Vorteile. So zeigen Studien, dass die Wahrscheinlichkeit Gewebe mit einem Roboter zu beschädigen mit haptischem Feedback niedriger ist als nur mit visueller Beobachtung [1]. Es wurde beobachtet, dass im Training zur Entfernung der Gallenblase 331 Fehler von 60 angehenden Chirurgen gemacht wurden. 55 Prozent dieser Fehler sind auf eine zu starke Kraftaufwendung zurückzuführen, die durch fehlendes Feedback entsteht [7].

Die Vorteile von haptischem Feedback wurden in mehreren Studien dargestellt [1,5,7], daher ist das Ziel dieser Arbeit verschiedene Ansätze zu vergleichen. Diese unterscheiden sich in der Sensorik, also wie die Kräfte vom System aufgenommen werden. Aber auch die

Aktorik, wie die aufgenommen Information an den Nutzer weitergegeben werden, unterscheiden sich. Unter diesen beiden Punkten sollen die Ansätze verglichen werden.

1.2 Struktur der Arbeit

Zunächst erfolgt eine Einführung in die genutzte Terminologie in dem Bereich des haptischen Feedbacks in der medizinischen Robotik. Danach geht es um die Herausforderungen, die bei einem solchen System bei der Entwicklung beachtet werden müssen. Anschließend werden unterschiedliche existierende Systeme im Kapitel Anwendungen analysiert. Zum Schluss erfolgt das Fazit.

2 Haptisches Feedback in der medizinischen Robotik

2.1 Terminologie

2.1.1 Somatische Sensibilität

Der Tastsinn gehört zu den fünf Sinnen des Menschen. Allerdings besteht die Wahrnehmung von Berührung aus der Zusammenführung von unterschiedlichen Rezeptoren. Dazu gehören die Merkel-Tastscheiben, die durch die mechanische Verformung der Haut gereizt werden. Außerdem können durch sie die Eindringtiefe und -dauer einer Berührung empfunden werden. Die Meissner-Körperchen registrieren die Eindrucksgeschwindigkeit einer Berührung. Die Vater-Pacini-Lamellenkörperchen reagieren auf Druck- und Vibrationsreize. Freie Nervenendigungen sind für Temperatur- und Schmerzreize sowie Juckreiz empfänglich. Diese Sinneswahrnehmungen lassen sich unter dem Begriff der somatischen Sensibilität zusammenfassen. Außerdem gibt es Muskelspindeln und Golgi-Sehnenorgane, mit denen die Dehnung der Muskeln und Sehnen und damit der benötigte Kraftaufwand wahrgenommen werden kann [9].

2.1.2 Haptische Interface Typen

Haptische Geräte werden durch den Somatischen Sinn, den sie stimulieren kategorisiert. Kinästhetische Geräte sollen Feedback für die auftretende Kraft beziehungsweise das

auftretende Drehmoment geben. Meist sorgen Elektromotoren dafür, dass die von den Sensoren aufgenommene Situation an den Menschen übertragen wird. Im taktilen Bereich wird unter anderem mit elektromagnetisch verstellbaren Nadeln, Formgedächtnismaterialien, piezoelektrischen Kristallen, pneumatischen Systemen und Heizungssystemen gearbeitet. Aufgrund der großen Vielfalt der Rezeptoren in diesem Bereich und der möglichen Umsetzungen, werden taktile Interfaces für einen bestimmten Zweck entwickelt. Auch sind taktile Displays komplizierter umzusetzen als kinästhetische Interfaces [1].

2.1.3 OP-Roboter: Teleoperation vs. Hands-On

In der Teleoperation besteht der OP-Roboter aus dem Master-Roboter und dem Slave-Roboter. Der Operateur kontrolliert mit Hilfe des Masters den Slave-Roboter, welcher sich am Patienten befindet. Der Raven II und der da Vinci sind Beispiele aus diesem Bereich der OP-Roboter. Diese Art wird häufig bei minimalinvasiven Operationen eingesetzt. Bei Hands-On Systemen befindet sich die Kontrolle direkt am operierenden Roboter. Ein Beispiel dafür ist das RIO System, welches bei Knieoperationen zum Einsatz kommt [1].

2.2 Herausforderungen

Eine der größten Herausforderungen ist die Stabilität des Systems. Nach der Kontrolltheorie beeinflussen Interaktionen mit dynamischer Kraft die Stabilität eines geschlossenen Regelkreises erheblich. So zeigten Versuche mit Robotern, dass haptisches Feedback liefern, dass bei Kontakt mit hartem Material das System instabil wird. In der Teleoperation ist mit weiteren Herausforderungen zu rechnen. Durch die Teilsysteme Operateur, Roboter und die Umgebung, die dynamisch miteinander agieren, entstehen weitere Möglichkeiten zur Instabilität. Die Feedbackschleife umfasst die Verzögerung der Verarbeitung und Reaktion des Operateurs. Dieser kann das System stabilisieren, aber auch ein

sonst stabiles System instabil werden lassen [1].

Eine weitere große Herausforderung bei der Implementierung von haptischem Feedback im Bereich der medizinischen Robotik ist die Aufzeichnung von haptischen Informationen im OP. Durch die Anforderungen an OP-Räume, die Größe und Robustheit von Robotern limitieren und die Schwierigkeit der Desinfizierbarkeit von Instrumenten, ist es schwer Sensoren zu entwickeln. Außerdem müssen die Sensoren selbst den Vorschriften der Europäischen Arzneimittel-Agentur oder der U. S. Food and Drug Administration (FDA) entsprechen. Dabei ist auch die Wirtschaftlichkeit zu beachten [1].

Um mögliche Vorteile von haptischen Systemen zu quantifizieren, muss eine klare Definition von objektiv messbaren Variablen vorhanden sein. Durch die implizite Natur der somatischen Sinne des Menschen ist dies eine große Herausforderung. Die unklare Definition einiger Ziele, die bei einer Operation erreicht werden sollen, erschwert die objektive Beurteilung weiter [1].

In der Teleoperation gibt es mehrere Faktoren, die eine Operation beeinflussen können. So ist hier eine quantitative Evaluation nur schwer möglich. Hier kann die Transparenz des Systems beobachtet werden. Transparenz beschreibt wie verständlich das Interface für den Benutzer ist. Ablenkungen, wie Vibrationen der Hardware oder Hitzeentwicklung durch elektrische Komponenten können die Transparenz beeinträchtigen. Die Transparenz eines haptischen Interfaces ist wie nah die Empfindung des Nutzers am Controller mit der Realität am Master übereinstimmt. So sollte sich zum Beispiel der Kontakt mit einem Knochen auf der Slave-Seite genauso anfühlen wie auf der Master-Seite [1].

Mit Transparenz allein lässt sich allerdings kein haptisches Interface evaluieren. So könnte es zum Beispiel ein System eine eingebaute Pfadfindung besitzen. Mit simulierten Kräften bleibt der Operateur so auf einem festgelegten Pfad. Das System ist also nicht

transparent, sondern will den Eingriff verbessern [1].

So muss also für jede Aufgabe eigene objektive Evaluierungsmöglichkeiten gefunden werden. Ebenfalls ist die Ausbildung der Probanden zu beachten, Laien erzielen in Studien regelmäßig größere Verbesserung mit haptischem Feedback als für medizinische Roboter geschultes Personal, welches eine bessere Hand-Augen-Koordination hat als der durchschnittliche Mensch [1].

2.3 Anwendung

2.3.1 Raven II

Der Raven II ist als Teleoperations-Plattform für Forschungszwecke entwickelt worden [2]. Das System besteht aus zwei Armen, siehe Abbildung 1, mit denen die austauschbaren Instrumente positioniert werden können. Insgesamt besitzt der Raven II sieben Freiheitsgrade, wobei ein Instrument vier besitzt und der Arm zur Positionierung drei Freiheitsgrade hat.

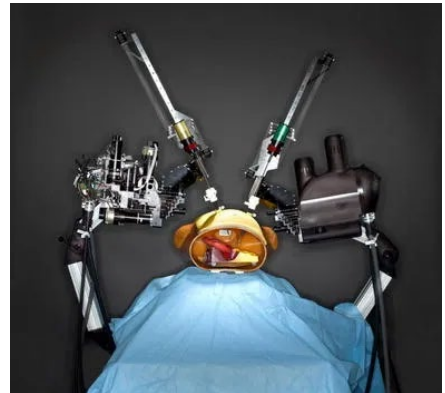


Abb. 1 Raven II Arme mit Instrumenten [5]

2.3.1.1 Sensorik

Das System ist als reine Forschungsplattform entwickelt worden, der Einsatz am Menschen ist nicht vorgesehen [2]. Die Position der operierenden Arme wird mit Encodern festgestellt. Der Computer errechnet aus den Positionswerten der Encoder ob der Roboter ein mechanisches Limit erreicht hat, oder ob es

eine Kollision zwischen den Armen gibt. Die auftretenden Kräfte werden dann über Phantom Omnis an den Nutzer weitergeleitet, um ihn bei Kollisionen zu warnen [5].



Abb. 2 Phantom Omni [5]

2.3.1.2 Aktorik

Kontrolliert werden die zwei Arme des Raven II mit zwei Phantom Omnis, siehe Abbildung 2. Diese Stift-artigen Eingabegeräte bieten haptisches Feedback. Die Controller können dabei eine Kraft von bis zu 3,3 Newton an den Nutzer weitergeben [5]. Die Greifer des Roboters werden über zwei Knöpfe an den Controllern geöffnet und geschlossen. Dabei wird kein haptisches Feedback weitergegeben. Der Operateur muss hier auf rein visuelle Kontrolle zurückgreifen. Bei Gewebe sind Deformationen und Farbänderungen wichtige Indikatoren [5].

2.3.2 MiroSurge

Der MiroSurge-Roboter wurde vom Deutschen Zentrum für Luft- und Raumfahrt (DLR) für Teleoperationen entwickelt. Das System besteht aus drei MIRO-Armen, siehe Abbildung 3, welche ebenfalls vom DLR entwickelt wurden. Zwei dieser Arme bieten, durch die eingesetzten MICA-Instrumente, Force Feedback. Ein dritter Arm ist mit zwei Kameras ausgerüstet, um ein Stereobild für einen 3D-Bildschirm, der vor dem Nutzer steht, zu erzeugen [5]. Die MICA-Instrumente besitzen drei Freiheitsgrade. Insgesamt besitzt ein MIRO-Arm mit dem MICA-Instrument sieben Freiheitsgrade [8].

2.3.2.1 Sensorik

An der Spitze des Instruments befindet sich ein vom DLR entwickelter F/T-Sensor, also ein Kraft-/Momentensensor, mit sieben Freiheitsgraden. So können die auftretende Kraft und das Drehmoment beim Bewegen des Instruments, als auch die entstehende Kraft beim Zugreifen gemessen werden. Der Sensor kann eine maximale Kraft von 10 Newton und eine minimale Kraft von 0,015 Newton messen [8].

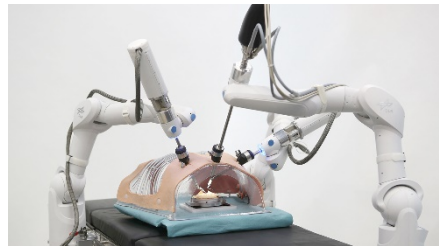


Abb. 3 MiroSurge [DLR]

2.3.2.2 Aktorik

Der MiroSurge wird mit dem von Force Dimension entwickelten haptischen Interface sigma.7 gesteuert. Dieses erlaubt eine sensible Manipulation und Palpation [8].

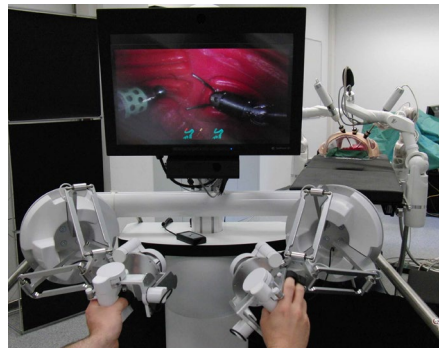


Abb. 4 sigma.7 steuert MiroSurge [5]

Dieses besteht aus zwei Manipulatoren, siehe Abbildung 4, mit jeweils sieben Freiheitsgraden. Jeder Manipulator besteht aus drei Teilen: Einer translationalen Basis, einem Gelenk zur Rotation und der Greiferkontrolle. Eine Besonderheit beim sigma.7 ist, dass die

Translation und Rotation voneinander getrennt sind. So kann der Operateur eine Rotationsbewegung durchführen ohne dass eine Translation passiert. Durch elektrische Motoren kann eine Kraft von bis zu 20 Newton erzeugt werden [8].

2.3.3 da Vinci

Der da Vinci Roboter von Intuitive Surgical ist eines der wenigen für die Teleoperation zugelassener Systeme. Der operierende Teil besteht aus mehreren Armen. Einer dieser Arme ist der Kameraarm, an welchem unterschiedliche endoskopische Kameras befestigt werden können. An den restlichen Armen, die Anzahl variiert je nach Modell, können eine große Vielfalt an medizinischen Instrumenten befestigt werden, siehe Abbildung 5. Die Arme besitzen mit den Instrumenten sieben Freiheitsgrade. Der da Vinci besitzt normalerweise kein haptisches Feedback [5].

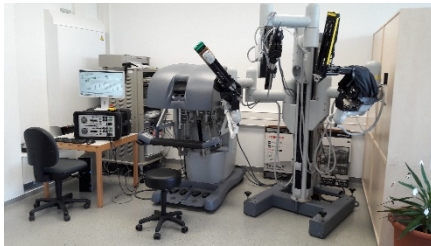


Abb. 5 da Vinci der Hochschule Reutlingen mit dVRK

Mit Hilfe des da Vinci Research Kit (dVRK), einer Plattform für die Forschung, konnte der da Vinci, statt mit der regulären Steuerung, mit einem sigma.7 gesteuert werden, siehe Abbildung 6. Das dVRK besteht aus einer Open-Source-Software und einer Hardwareplattform, mit welchen da Vincis der ersten Generation gesteuert werden können. Dazu werden die Arme über die Hardware an einen Computer angeschlossen. Nun können durch die Software Daten ausgelesen und Motoren gesteuert werden. So konnten Saracino et al. [4] den da Vinci mit dem sigma.7 verbinden.

2.3.3.1 Sensorik und Aktorik

Die auftretenden Kräften wurden dabei durch den Stromverbrauch der Elektromotoren,

welche die Arme steuern, berechnet. Diese wurden dann für den sigma.7 skaliert, so entsprechen 30 Newton am Slave, 7,5 Newton am Controller. Beim Greifen wurde ein Kraftaufwand von 0,12 Newton am Instrument auf 1,2 Newton am sigma.7 hochskaliert. Die Werte am Instrument wurden durch vorhergehende Experimente festgestellt und sind geschätzte Maximalwerte. Da der sigma.7 aber nur bis zu 20 Newton erzeugen kann, wurde auf die genannten Werte skaliert [4].

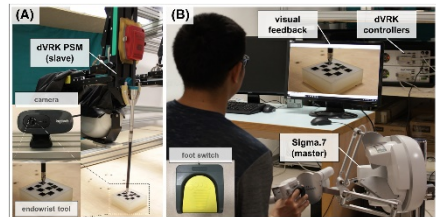


Abb. 6 dVRK-Aufbau mit sigma.7 [4]

In der folgenden Studie wurde dieser Ansatz mit Erfolg getestet. Als Probanden dienten dabei Menschen ohne chirurgische Ausbildung, Assistenzärzte und erfahrene Chirurgen. Innerhalb der Studie mussten mehrere Aufgaben aus drei Bereichen, jeweils mit und ohne haptisches Feedback, absolviert werden. Dazu zählen Greifen, Palpation, also das Abtasten von Gewebe und Schnitte. Obwohl objektiv keine große Verbesserung festzustellen war, sagten die Ärzte, dass das haptische Feedback sie in ihrer Aufgabe unterstützt habe. Der einzige Kritikpunkt der teilnehmenden Ärzte war das teilweise als zu gering empfundene Force Feedback (Kraft-rückkopplung). Dies könnte, laut den Autoren, auch dazu geführt haben, dass objektiv keine Verbesserung bei angestelltem haptischem Feedback zu messen war [4].

Insgesamt ist zu sagen, dass dieser Ansatz sehr interessant ist, da keine Sensoren eingebaut werden müssen, welche wie in Kapitel 2.2 beschrieben, viele Herausforderungen mit sich bringen.

2.3.4 Acrobot

Der Acrobot ist ein Hands-On System, welches bei der Einsetzung eines künstlichen Kniegelenks benutzt wird [5]. Um dies zu erreichen müssen die umgebenden Knochen in eine bestimmte Form gebracht werden. Damit eine optimale Passung erreicht werden kann, muss mit hoher Präzision gearbeitet werden. Nur so kann sichergestellt werden, dass der Patient nach dem Eingriff beschwerdefrei laufen kann. In einer normalen Operation benutzt der Chirurg dafür eine oszillierende Säge und eine Vielzahl von Fixierungsmethoden, um eine präzise Entfernung zu ermöglichen. Dies ist mit dieser Methode allerdings sehr schwer zu erreichen. Aus diesem Grund wurde Acrobot entwickelt [3].

Vor einer Operation mit dem Acrobot muss das Gebiet, das entfernt werden soll, festgelegt werden. Dieser Teil der Planung ist aufwendiger als bei der normalen Operation. Zunächst muss ein CT-Bild des Beines angefertigt werden. Dieses wird dann in eine speziell entwickelte Software geladen. Diese berechnet mit Hilfe der CT-Bilder ein 3D-Modell des Beins. Der Chirurg kann dann das künstliche Kniegelenk interaktiv hinzufügen. Wenn dieses an seinem Platz ist berechnet die Software den Teil des Knochens, der entfernt werden muss. Während der Operation muss der Patient fixiert werden, um ein Verrutschen zu vermeiden. Wenn der Patient fixiert ist, muss eine Registrierung durchgeführt werden. Das erstellte virtuelle Modell muss mit der realen Welt übereinstimmen. Dies passiert durch einen iterativ closest point Algorithmus (ICP). Der Chirurg wählt zunächst vier Landmarken, mit diesen kann der Algorithmus eine erste Einschätzung berechnen. Um die Registrierung zu verfeinern werden weitere 20 bis 30 Punkte auf den Knochen zufällig ausgewählt [3].

2.3.4.1 Aktorik

Bei der OP kann der Chirurg dann nur den vorher berechneten Teil entfernen. Grenzen werden durch eine immer stärker werdende gegen gelenkte Kraft verdeutlicht, dies wird als active constraint control bezeichnet.

Durch die graduelle Verstärkung wird die Stabilität des Systems gewährleistet. Der Acrobot hat drei orthogonale Bewegungsachsen. Der Roboter kann sich jeweils in Rahmen von -30 bis 30° in der Gier- und Nickachse bewegen, siehe Abbildung 7. Außerdem kann die Fräse von der Anfangslänge von 30 Zentimetern auf 50 Zentimeter ausgefahren werden [3].

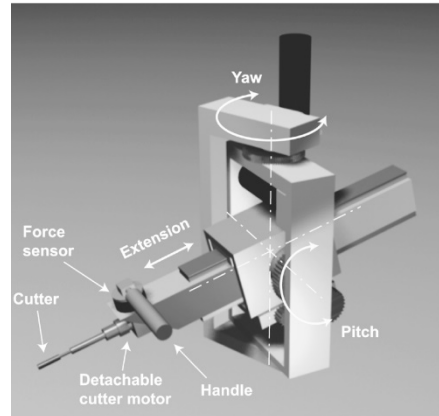


Abb. 7 Acrobot Schematische Darstellung [3]

2.3.4.2 Sensorik

Der Acrobot ist relativ sicher, da diese Bewegungseinschränkungen und die schwachen Elektromotoren dafür sorgen, dass möglicher Schaden minimal gehalten wird. Der Operateur bedient, über den Griff, selbst die Fräse, der Acrobot bietet die benötigte Stabilität. So kann der Chirurg selbst die Struktur des Knochens fühlen. Am Griff befindet sich ein sechssachsiges F/T-Sensor. Die Information dieses Sensors werden für die active constraint control benötigt [3].

Der Acrobot wurde am Imperial College of Science, Technology and Medicine in London entwickelt. Es folgte die Ausgründung. Schließlich wurde die Acrobot Company von Stanmore Implants aufgekauft, welche wiederum von Stryker gekauft wurde. Die Technologie des Acrobot wird in der MAKO-Serie weiterverwendet [5].

2.3.5 neuroArm

Der neuroArm wurde an der University of Calgary für neurologischen Operationen entwickelt. Bei diesem Roboter handelt es sich um ein Teleoperationssystem, welches für den MRT-Einsatz optimiert wurde [5]. An der Konsole des Operators können so unter anderem aktuelle MRT-Bilder angezeigt werden.



Abb. 8 Konsole mit Omnis [5]

Die Konsole besitzt mehrere Bildschirme, sodass ein breites Feld an nützlichen Informationen und radiologischen Bildern angezeigt werden können, siehe Abbildung 8 [6].

2.3.5.1 Sensorik

Der neuroArm besitzt zwei Arme mit austauschbaren Instrumenten, welche drei Freiheitsgrade besitzen. Insgesamt hat jeder Arm mit Instrument sieben Freiheitsgrade. Jeder Arm besitzt zwei Sensoren, mit denen die aufgewendete Kraft an die Konsole geschickt wird [6]. Da der Roboter für den MRT-Einsatz entwickelt wurde, konnte kein normaler F/T-Sensor eingebaut werden. Daher benutzt der neuroArm optische Kraft-Sensoren. Diese können durch eine Methode aus der Spannungsoptik die Spannungsverteilung, mit Hilfe von polarisiertem Licht, in den Bauteilen erkennen [6].

2.3.5.2 Aktorik

Die Konsole bestand zunächst aus zwei modifizierten PHANTOM Omnis. Später wurden diese durch einen sigma.7 Controller ersetzt, da diese auf sieben Freiheitsgraden haptisches Feedback erlaubt. Eine Besonderheit des neuroArm ist das integrierte OP-Mikroskop, siehe Abbildung 9 [6].



Abb. 9 neuroArm im Einsatz [5]

Hier kann ein Assistenzarzt direkt den Eingriff beobachten und dem Operateur durch ein Kommunikationssystem Hinweise geben. Dies ist während der gesamten Operation möglich. Der neuroArm ist nicht kommerziell erwerbbar, allerdings wurden schon erfolgreiche Studien an Menschen durchgeführt [6].

2.4 Vergleich Sensorik

Teleoperation				Hands-On	
Mit Sensor		Ohne Sensor		Sensor und Mensch	
MiroSurge/ neuroArm		da Vinci/ Raven II		Acrobot	
Funktionsweise: - Messung der Kraft über F/T-Sensoren ODER - Messung der Kraft über optische Kraftsensoren -> Spannungsverteilung mit polarisiertem Licht erkennen -> Geeignet für MRTs		Funktionsweise: - Berechnen der Kraft über Encoderwerte -> Nutzen des Stromverbrauchs ODER - Positionsbestimmung, um bei Kollision der Arme zu warnen		Funktionsweise: - Der Mensch fühlt über mechanische Verbindungen direkt die auftretenden Kräfte - Sensor wird für die Einhaltung von Grenzen benötigt (active constraint control)	
Pro	Contra	Pro	Contra	Pro	Contra
- Genaue Messung - Dedi-zierte für haptisches Feedback entwickelte Systeme	- Teuer - Schwer zu desinfizieren - Schwer zu verbauen - Zulassung schwer zu erreichen -> Stabilität muss gewährleistet sein	- Billig - Bzw. schon in OP-Robotern verbaut -> Existierende Roboter können aufgerüstet werden - Müssen nicht desinfiziert werden	- Ungenau	- Keine Verfälschung der Kräfte - Stabiles System - Medizinische Zulassung einfacher zu erreichen als bei Teleoperations-systemen -> Erfahrener Chirurg fühlt die Auswirkungen direkt	- Intensive Planung nötig

2.5 Vergleich Aktorik

Teleoperation				Hands-On	
sigma.7 MiroSurge/neuroArm/ da Vinci		Phantom Omni Raven II/ neuroArm(zunächst)		Acrobot und Mensch	
Funktionsweise: - 2 Manipulatoren mit: - 7 Freiheitsgrade mit haptischem Feedback bis zu 20 Newton - Translatationaler Basis, Gelenk zur Rotation, Greiferkontrolle		Funktionsweise: - Stiftartiges Eingabegerät - Haptisches Feedback in 3 Freiheitsgraden bis zu 3,3 Newton - Greiferkontrolle über zwei Knöpfe		Funktionsweise: - Mensch steuert das Werkzeug über einen Griff am Roboter direkt - Haptisches Feedback signalisiert über Stärke der Kraft das Erreichen von vorher festgelegten Grenzen -> 3D Bild aus CT-Bildern -> Registrierung vor der OP, damit Modell und Realität übereinstimmen - Eingeschränkte Bewegungsachsen	
Pro - Für OP besonders gut geeignet - Ergonomisch - Haptisches Feedback in allen Freiheitsgraden	Contra - Schwerer zu beschaffen	Pro - Intuitiv durch Form - Leicht zu beschaffen	Contra - Kein haptisches Feedback beim Greifen - Für OP bedingt geeignet - Stärke des haptischen Feedbacks ungenügend	Pro - Sehr sicheres und stabiles System - Stärken von Roboter und Mensch werden kombiniert -> Roboter: Stabilität und Genauigkeit -> Mensch: Gefühl und Erfahrung	Contra - Längere Planungsphase

3 Fazit

Diese Arbeit zeigt, dass die Verbreitung von haptischem Feedback in der medizinischen Robotik noch gering ist. Alle vorgestellten Anwendungen sind bis auf den Acrobot Forschungsprojekte und keine zugelassenen Produkte. Jedoch ist der positive Faktor von haptischem Feedback durch mehrere Studien bewiesen worden. So wird weiterhin daran gearbeitet die in Kapitel 2.2 genannten Herausforderungen zu meistern und ein System mit haptischem Feedback zu entwickeln, dass alle medizinischen Standards erfüllt.

Die in Kapitel 2.3 vorgestellten Systeme haben unterschiedliche Ansätze des haptischen Feedbacks gewählt. Der Raven II und der modifizierte da Vinci benutzen die Werte der Encoder, um auftretende Kräfte zu berechnen. MiroSurge und neuroArm haben dedizierte Sensoren, um die Kräfte zu messen. Der Acrobot als Hands-On-System ist hier eine Besonderheit. Jeder dieser Ansätze hat unterschiedliche Vor- und Nachteile. Die Berechnung von auftretenden Kräften durch Encoderwerte benötigt keine teuren Sensoren, die schwer zu desinfizieren sind. Allerdings können die Ergebnisse sehr ungenau sein. Dies wird durch dedizierte Sensoren verhindert. Diese bieten eine hohe Messgenauigkeit. Sensoren sind allerdings teurer in der Anschaffung, schwer zu desinfizieren und müssen klein genug sein, um in OP-Besteck verbaut werden zu können. Bei der direkten Übertragung des haptischen Feedbacks, wie beim Acrobot entsteht dieses Problem nicht. Der Roboter liefert die benötigte Stabilität und Genauigkeit, während der Mensch die auftretenden Kräfte ohne Verfälschung spüren kann. Durch diese Maßnahmen ist ein stabiles System garantiert. So ist der Acrobot mit seinen Nachfolgern auch der einzige medizinisch zugelassene Roboter mit haptischem Feedback in dieser Arbeit.

In den vorgestellten Teleoperationssystemen werden entweder Phantom Omnis oder ein sigma.7-Controller eingesetzt. Der stiftartige Phantom Omni ist eher ungeeignet für den medizinischen Einsatz, da beim Zugreifen

kein haptisches Feedback gegeben werden kann. Aber auch in den restlichen Freiheitsgraden kann maximal nur eine Kraft von 3,3 Newton wiedergegeben werden. Der sigma.7 hingegen wurde für diese Art der Anwendung entwickelt. Er bietet haptisches Feedback für alle sieben Freiheitsgrade und kann dabei eine Kraft von bis zu 20 Newton an den Benutzer weiterleiten. Aus diesen Gründen hat sich der sigma.7 zum Standard-Controller für Teleoperationsroboter mit haptischem Feedback entwickelt [5].

4 Literaturverzeichnis

- [1] N. Enayati, E. De Momi, G. Ferrigno: Haptics in Robot-Assisted Surgery: Challenges and Benefits, IEEE Reviews in Biomedical Engineering, Vol. 9, S. 49-60, 2016
- [2] B. Hannaford, J. Rosen, D. W. Friedmann, H. King, P. Roan, L. Cheng, D. Glozman, J. Ma, S. N. Kosari, L. White. Raven-II: An Open Platform for Surgical Robotics Research. IEEE Transactions on Biomedical Engineering, Vol. 60, No. 4, 2013
- [3] M. Jakopcic, S. J. Harris, F. R. y Baena, P. Gomes, B. L. Davies. The Acrobot system for total knee replacement. Industrial Robot: An International Journal, Vol. 30, No. 1, 2003
- [4] A. Saracino, A. Deguet, F. Staderini, M. N. Boushaki, F. Cianchi, A. Menclasi, E. Sinibaldi. Haptic Feedback in the da Vinci Research Kit (dVRK): A user study based on grasping, palpation and incision tasks. Wiley. The International Journal of Medical Robotics and Computer Assisted Surgery. 2019
- [5] A. Simorov, R. S. Otte, C. M. Kopietz, D. Oleynikov. Review of surgical robotics user interface: what is the best way to control robotic surgery? Springer Science+Business Media, 2012

- [6] G. R. Sutherland, S. Wolfsberger, S. Lama, K. Zarei-nia. The Evolution of neuroArm. *Neurosurgery*, Vol. 51, No. 51, 2013
- [7] B. Tang, G. B. Hanna, A. Cuschieri: Analysis of errors enacted by surgical trainees during skills training courses, *Surgery*, vol. 138, no. 1, S. 14–20, 2005
- [8] A. Tobergte, P. Helmer, U. Hagn, P. Rouiller, S. Thielmann, S. Grange, A. Albu-Schäffer, F. Conti, G. Hirzinger. The sigma.7 haptic interface for Miro-Surge: A new bi-manual surgical console. *IEEE/RSJ International Conference on Robots and Systems*, San Francisco, CA, USA, 2011.
- [9] C. Zalpour: *Anatomie Physiologie für die Physiotherapie*, 4. Auflage; Elsevier GmbH, München, 2016

Disease Detection and Health Monitoring Through Breath Analysis

Marc Viertel

Reutlingen University

Marc.Viertel@Student.Reutlingen-University.DE

Abstract

The field of breath analysis has developed to be of growing interest in medical diagnosis and patient monitoring. The main advantages are that it's noninvasive, painless and repeatable in flexible cycles. Even though breath analysis is being researched for a couple of decades there are still many unanswered questions. Human breath contains volatile organic compounds which are emitted from inside the body. Some of these compounds can be assigned to specific sources, such as inflammation or cancer, but also to non-health related origins. This paper gives an overview of breath analysis for the purpose of disease diagnosis and health monitoring. Therefore, literature regarding breath analysis in the medical field has been analyzed, from its early stages to the present. As a result, this paper gives an outline of the topic of breath analysis.

Keywords

Breath Analysis, Electronic nose, Volatile organic compounds, VOCs, Disease diagnostic

CR-Categorys

- Applied computing~Life and medical sciences~Bioinformatics
- Applied computing~Life and medical sciences~Health informatics
- Applied computing~Life and medical sciences~Metabolomics / metabonomics

1 Introduction

During the past decades there has been a growing interest in using breath analysis in the medical field. This can be clearly seen when searching in the PubMed database for the terms 'breath analysis' and 'disease' as shown in Figure 1. Breath analysis as a diagnostic tool provides several advantages over other procedures like blood testing or biopsy. Breath analysis is noninvasive, painless, real-time capable and basically can be performed as often as required. Human breath contains several volatile organic compounds (VOCs) which are of versatile origin either endogenous¹ or exogenous² [14]. For the purpose of medical diagnosis only the VOCs of endogenous origin are of relevance. The range of different VOCs contained in human breath is wide. Costello et al. state over 800 VOCs related to human breath, from which around 200 are clearly identified [9]. Aside from human breath, VOCs are also emitted by other sources like skin, urine, blood, saliva, human-breast milk and feces [9]. Some of the VOCs emitted in human breath can be assigned to specific exogenous behavior such

Academic supervisor: Prof. Dr. rer. medic. Dipl.-Inform.
Christian Thies
Reutlingen University
Christian.Thies@Reutlingen-University.DE

Wissenschaftliche Vertiefungskonferenz
20. Mai 2020, Reutlingen University
Copyright 2020 MarcViertel

¹ Originating from inside the body

² Environmental/external influences on the body

as smoking. Other VOC concentrations are influenced in case of diseases, some VOCs even only exist in human breath of unhealthy patients [1]. One major problem of breath analysis is making it comparable. The proportion of different VOCs in human breath differs even when compared between healthy patients [13].

Aside from analyzing breath with technical instruments it is also possible to conclude a diagnose by simply using the olfactory³ sense. Many diseases come with a characteristic smell. Hayden proposed an overview with unusual breath odors which can be associated to diseases or ingestions. A sweet smell can be an indicator for ketoacidosis⁴ which is related to diabetes mellitus, for example [4].

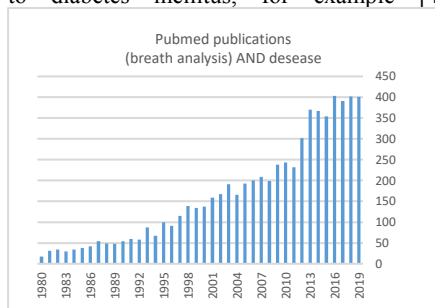


Figure 1: Publications listed on PubMed since 1980 till today for the terms 'breath analysis' and 'disease'. The growing interest can be clearly seen.

1.1 Approach and methodology

This work is based on a literature research using the terms 'breath analysis', 'disease detection', 'Electronic nose', and 'volatile organic compounds'. There were early works published in the mid of the 1960s [18] to some recent works published just recently in 2020 [6].

1.2 Structure of work

This paper is structured into some basics on breath analysis in Chapter 2 followed by some

insights into the technical part of VOC measurement in Chapter 3. Medical application of breath analysis is discussed in Chapter 4, which shows the diagnosis of patient diseases. Finally, Chapter 5 concludes on the state of the art of breath analysis and closes with an outlook for future developments.

2 Breath Analysis

Unlike diagnostic procedures as blood test or biopsy, breath analysis strikes to be simple and breath itself is easy to acquire. But still, breath sampling needs attention, as there are differences in the acquisition of breath, which can significantly influence the results. Basically, there are two different ways breath sampling is performed [10]:

- Mixed expiratory sampling
- Alveolar sampling

The two principles are shown in Figure 2. A rise in CO₂ in the breath means that the air comes from inside the lung. Phase I in the Figure refers to air which is called *dead space* gas. *Dead space* gas does not take part in the gas exchange inside the lungs and mostly stays in the upper airways. Therefore, this proportion of the exhaled breath does usually not contain endogenous VOCs. The breath in phase II is a mix of dead space gas and gas originating from the alveolar cells of the lung. Phase 3 should contain only gas from the alveolar cells and therefore has the highest concentration of endogenous VOCs. When phase III ended and the alveolar cells are empty, another inspiratory phase (IV) begins. For mixed sampling the gas from phase I-III is sampled. For alveolar sampling only the gas originating from the alveolar cells in phase III is sampled [10]. When taking a mixed expiratory sample, the dead space gas and the environmental influence on the sample has to be considered. Hence, the sample can be corrected, for example, by subtracting a sample of the environmental gas. Another way

³ Sense of smell

⁴ Complication caused by absolute insulin deficiency

to eliminate environmental influence, is to let the patient breath pure air for a few minutes before the sample process for breath analysis is carried out [16]. Also, condensation of water vapor from exhaled breath has to be taken into account while taking a sample. If condensation occurs, exogenous compounds may be attracted, leading to an erroneous measurement [7].

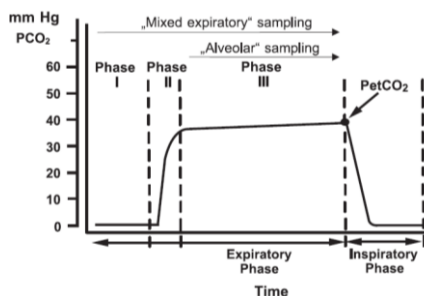


Figure 2: The diagram shows a capnogram which measures CO₂ concentration in breath gas. As CO₂ rises after lung gas exchange, it is good to distinguish between dead space gas which did not take part in lung gas exchange and alveolar gas. Phase I is dead space gas which does not contain CO₂ and no endogenous VOCs. Phase II represents a mix of dead space gas and alveolar gas. Phase III contains only gas from alveolar cells and therefore the highest concentration of endogenous VOCs. Phase IV shows the next inspiratory cycle. Phase I-III is used for mixed breath sampling. Phase III is used for alveolar sampling [10].

The main indicators for medical diagnosis by breath analysis are the VOCs mentioned above. Some of the endogenous VOCs can serve as biomarkers which indicate specific diseases. Acetone, for example, can be an indicator for diabetes. The arterial blood contains acetone which is diffused by the lungs, therefore it can be measured by breath analysis [18].

Another metabolism which can be measured is muscle exertion. The hydrocarbon

2-methyl-1,3-butadiene, also known as Isoprene, can be measured in exhaled breath. Isoprene is thought to be stored in body tissue as muscles [1, 7]. Due to muscle exertion, a proportion of Isoprene is released from muscles and gets into the blood circulation. After some time, the Isoprene saturated blood reaches the lungs where the gas exchange through the alveolar cells is carried out. The Isoprene evaporates into the exhaled gas and thus can be measured by breath sampling. The whole process is shown in **Figure 3**.

When performing breath analysis studies, the variability of breath composition must be taken into account. For example, some VOC biomarkers are age- or gender dependent. In cross-sectional studies, in which a healthy control group is compared with a diseased group, the quantitative difference of the measured biomarkers should be large enough to differentiate. This problem is smaller when longitudinal studies are conducted in which, for example, a disease or medication is monitored for single- or a group of patients [10].

3 Measurement of volatile organic compounds in breath

After a breath sample is collected it needs to be analyzed. In contrast to other diagnostic tools which need to be analyzed inside a laboratory, breath can be analyzed in real-time [1]. This is one major advantage of breath analysis, because the medical staff can apply breath analysis on demand in an instant manner and as often as necessary. Mostly the breath sample can't be analyzed directly. This is because most substances contained in human breath only exist in concentrations between ppb⁵ and ppt⁶. To bypass this problem, the sample is pre-concentrated before being processed any further. The process of pre-concentration is not described in detail here, because of its own complexity in the field of chemistry and physics. For further information the paper published by Mochalski et al. is recommended [11]. In a nutshell pre-

⁵ Parts per billion

⁶ Parts per trillion

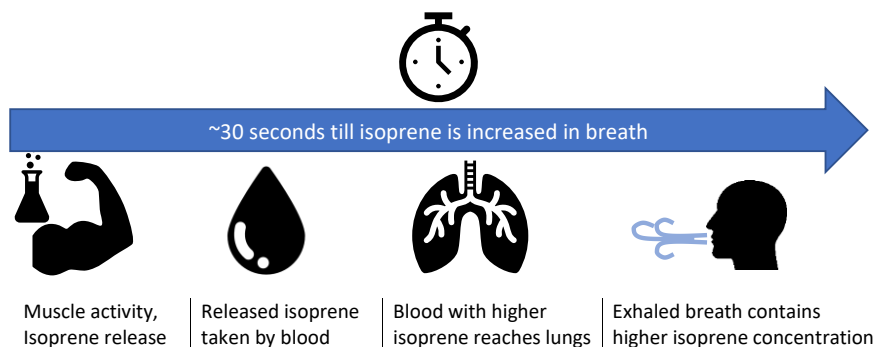


Figure 3: Metabolism of isoprene contained in muscle tissue. During muscle exertion the muscle perfusion is increased and therefore a higher concentration of isoprene is taken up by the blood. The blood reaches the lungs after some time, where the gas exchange through the alveolar cells is carried out. Finally, the isoprene as VOC is measurable in the breath. A short time after the muscle exertion started, the isoprene concentration decreases again as there is no endless reservoir.

concentration of breath samples helps to highlight traces of compounds which otherwise may be lost in the original sample [11].

After pre-concentration is finished, the now concentrated sample can be analyzed. This can be accomplished by various analytical techniques, of which gas chromatography-mass spectrometry (GC-MS) is the historically most used technique. In principle it splits up the gas in its single components which then can be measured [10]. More recent analytical techniques based on spectrometry developed in the past 30 years are:

- Proton-transfer-reaction mass spectrometry (PTR-MS) [21]
- Selected ion flow tube mass spectrometry (SIFT-MS) [20]
- Ion mobility spectrometry (IMS) [19]
- Laser spectrometry [17]

All of the four mentioned analytical techniques are able to do real-time measurements [1].

Aside from spectrometry there are other possibilities for measurement, for example, the field of conductometry. Electronic parts are used to measure gas which comes in contact with the sensor. Depending on the gas composition the measured conductivity changes [3]. One device that makes use of gas sensors is the *Aeonose*⁷. This device implements three metal oxide sensors for breath gas analysis. Van de Goor et al. showed in a study that it is possible to differentiate between patients with and without lung cancer [23].

4 Breath analysis for medical applications

As mentioned before, breath analysis offers a set of advantages towards classic diagnostical procedures. It is easy to acquire, painless, non-invasive and real-time capable. Breath analysis can even be applied during surgery or sleep [1]. Accordingly, breath analysis is expected to be introduced into clinical care to help diagnose diseases and monitor patients health or pharmacological effects [10].

⁷ <https://www.enose.nl/products/aeonose/>

Breath analysis can be used for medical diagnosis as shown by many researchers. The bacteria *Helicobacter pylori* which infects the stomach of patients and is known to cause gastric diseases like gastric cancer can be diagnosed using breath analysis. The patient needs to ingest special labeled urea. This means, that the urea is labelled with an uncommon isotope. After a short waiting period of around 10 minutes a breath sample is taken and compared with a beforehand sampled measurement. If these labeled compounds are present in the sample, the patient is infected with the bacteria. This works due to the fact, that only the bacteria are able to split the urea and therefore labelled CO₂ is exhaled. As a result, if the bacteria is present for a patient, the clinical staff knows that there is an increased chance of gastric disease [15].

In 2019, Kort et al. proposed methods to improve lung cancer diagnosis by combining the data acquired by breath analysis, with clinical parameters from each patient. The breath samples were taken with the before mentioned *Aeonose* device in Chapter 3. As a result, they recommended adding clinical information to the classification value calculated by the *Aeonose*. With their proposed model they were able to achieve a sensitivity of 95.7%. This model can be used to conduct some early lung cancer screenings and in combination with the common computer tomography for diagnosis to reduce false-positive cases which occasionally occur [8].

With regard to the current situation of the COVID-19 pandemic, Khoubnasabjafari et al. just proposed a method for early diagnosis [6]. They propose to analyze the exhaled breath condensate (EBC), which is a condensed form of lung lining fluid contained in human breath. This approach is not relying on VOCs, but EBC still contains traces of bacteria and viruses [6, 24].

The EU Project *VOGAS*⁸ strives to achieve gastric cancer screening based on breath sampling. In a project related publication, Jaeschke et al. propose a modular system based on several analogue and digital metal oxide sensors. The device is designed to capture the alveolar part of the exhaled breath. They tested the device with low concentrations of VOCs like isopropanol, isoprene, n-pentane and acetone. After applying principal component analysis, the VOCs could be clearly differentiated [5]. The device proposed for breath sampling in the *VOGAS* project is shown in Figure 4. There is a mouth piece on the left side through which the patient exhales, a sampling tube and the three sensor modules containing the different gas sensors [5].

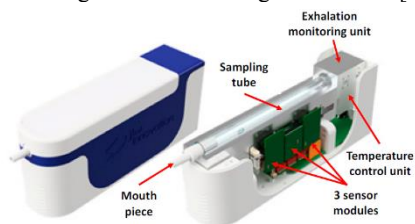


Figure 4: Modular breath sampling device proposed due to the VOGAS project. On the left side is the mouthpiece through which the patient breathes. The gas is lead through the sampling tube and analyzed by the sensor modules [5].

In other studies a wide spread of diseases are diagnosed through breath analysis like chronic obstructive pulmonary disease (COPD) [22], asthma [12], allograft⁹ rejection [12], liver disease [2] and others.

5 Conclusion & Outlook

This work shows, that there is intense research ongoing in the field of breath analysis. Breath analysis offers many advantages to clinical staff as well as to the patient in care. It can be easily applied, even during sleep. Also, it is often cheaper to apply than other diagnostic procedures. Despite the intense research since

⁸<https://www.vogas.eu>

Last accessed on 26.04.2020

⁹ A transplant from a genetically not identical donor

a few decades and many benefits, not many applications have found their way into clinical use. This may be owed to many technical variables which need to be controlled and standardized. Also, breath analysis seems not to be reliable enough in some applications to serve as stand-alone procedure for diagnosis. Still, it may be combined easily with other diagnostic tools to increase the certainty of a diagnosis given. It is to expect that the research in human breath analysis is conducted further on and more applications are brought to market. Devices as the *Aeonose* or the breath sampler from the *VOGAS* project let assume, that these devices are also able to perform measurements for the purpose of several different disease diagnostics. The diagnostic tool of breath analysis will gain more and more of importance through its whole set of benefits towards already applied procedures as biopsy. The success of many disease treatments depends on early recognition of these, especially for cancer in any form. A futuristic vision could be some kind of screening device, which samples human breath for a few seconds and outputs directly the status for a list of common widespread diseases. To achieve visions as the mentioned, still a lot of challenges need to be solved. A lot of these challenges are related to neural networks, signal processing, algorithms as well as improved understanding of VOC origins.

6 References

- [1] Anton Amann, Wolfram Miekisch, Jochen Schubert, Bogusław Buszewski, Tomasz Ligor, Tadeusz Jezierski, Joachim Pleil, and Terence Risby. 2014. Analysis of exhaled breath for disease detection. *Annual review of analytical chemistry (Palo Alto, Calif.)* 7, 455–482. DOI: <https://doi.org/10.1146/annurev-anchem-071213-020043>.
- [2] S. Chen, L. Zieve, and V. Mahadevan. 1970. Mercaptans and dimethyl sulfide in the breath of patients with cirrhosis of the liver. Effect of feeding methionine. *The Journal of laboratory and clinical medicine* 75, 4, 628–635.
- [3] Ananya Dey. 2018. Semiconductor metal oxide gas sensors: A review. *Materials Science and Engineering: B* 229, 206–217. DOI: <https://doi.org/10.1016/j.mseb.2017.12.036>.
- [4] G. F. Hayden. 1980. Olfactory diagnosis in medicine. *Postgraduate medicine* 67, 4, 110–5, 118. DOI: <https://doi.org/10.1080/00325481.1980.11715427>.
- [5] Carsten Jaeschke, Oriol Gonzalez, Marta Padilla, Kaylen Richardson, Johannes Glöckler, Jan Mitrovics, and Boris Mizaikoff. 2019. A Novel Modular System for Breath Analysis Using Temperature Modulated MOX Sensors. *Proceedings* 14, 1, 49. DOI: <https://doi.org/10.3390/proceedings2019014049>.
- [6] Maryam Khoubnasabjafari, Vahid Jouyban-Gharamaleki, Reza Ghanbari, and Abolghasem Jouyban. 2020. Exhaled breath condensate as a potential specimen for diagnosing COVID-19. *Bioanalysis*. DOI: <https://doi.org/10.4155/bio-2020-0083>.
- [7] J. King, A. Kupferthaler, K. Unterkofler, H. Koc, S. Teschl, G. Teschl, W. Miekisch, J. Schubert, H. Hinterhuber, and A. Amann. 2009. Isoprene and acetone concentration profiles during exercise on an ergometer. *Journal of breath research* 3, 2, 27006. DOI: <https://doi.org/10.1088/1752-7155/3/2/027006>.
- [8] Sharina Kort, Marjolein Brusse-Keizer, Jan W. Gerritsen, Hugo Schouwink, Emanuel Citgez, Frans de Jongh, Jan van der Maten, Suzy Samii, Marco van den Bogart, and Job van der Palen. 2020. Improving lung cancer diagnosis by combining exhaled-breath data and clinical

- parameters. *ERJ open research* 6, 1. DOI: <https://doi.org/10.1183/23120541.00221-2019>.
- [9] B. de Lacy Costello, A. Amann, H. Al-Kateb, C. Flynn, W. Filipiak, T. Khalid, D. Osborne, and N. M. Ratcliffe. 2014. A review of the volatiles from the healthy human body. *Journal of breath research* 8, 1, 14001. DOI: <https://doi.org/10.1088/1752-7155/8/1/014001>.
- [10] Wolfram Miekisch, Jochen K. Schubert, and Gabriele F. E. Noeldge-Schomburg. 2004. Diagnostic potential of breath analysis--focus on volatile organic compounds. *Clinica chimica acta; international journal of clinical chemistry* 347, 1-2, 25-39. DOI: <https://doi.org/10.1016/j.cccn.2004.04.023>.
- [11] Paweł Mochalski, Beata Wzorek, Ireneusz Sliwka, and Anton Amann. 2009. Improved pre-concentration and detection methods for volatile sulphur breath constituents. *Journal of chromatography. B, Analytical technologies in the biomedical and life sciences* 877, 20-21, 1856-1866. DOI: <https://doi.org/10.1016/j.jchromb.2009.05.010>.
- [12] C. O. Olopade, M. Zakkar, W. I. Swedler, and I. Rubinstein. 1997. Exhaled pentane levels in acute asthma. *Chest* 111, 4, 862-865. DOI: <https://doi.org/10.1378/chest.111.4.862>.
- [13] Michael Phillips, Jolanta Herrera, Sunithi Krishnan, Mooena Zain, Joel Greenberg, and Renee N. Cataneo. 1999. Variation in volatile organic compounds in the breath of normal humans. *Journal of Chromatography B: Biomedical Sciences and Applications* 729, 1-2, 75-88. DOI: [https://doi.org/10.1016/S0378-4347\(99\)00127-9](https://doi.org/10.1016/S0378-4347(99)00127-9).
- [14] J. D. Pleil and A. B. Lindstrom. 1997. Exhaled human breath measurement method for assessing exposure to halogenated volatile organic compounds. *Clinical Chemistry* 43, 5, 723-730.
- [15] R. E. Pounder and D. Ng. 1995. The prevalence of *Helicobacter pylori* infection in different countries. *Alimentary pharmacology & therapeutics* 9 Suppl 2, 33-39.
- [16] Terence H. Risby and Shelley S. Sehnert. 1999. Clinical application of breath biomarkers of oxidative stress status. *Free Radical Biology and Medicine* 27, 11-12, 1182-1192. DOI: [https://doi.org/10.1016/S0891-5849\(99\)00212-9](https://doi.org/10.1016/S0891-5849(99)00212-9).
- [17] Terence H. Risby and Frank K. Tittel. 2010. Current status of midinfrared quantum and interband cascade lasers for clinical breath analysis. *Opt. Eng* 49, 11, 111123. DOI: <https://doi.org/10.1117/1.3498768>.
- [18] Gösta Rooth and Stig Östenson. 1966. ACETONE IN ALVEOLAR AIR, AND THE CONTROL OF DIABETES. *The Lancet* 288, 7473, 1102-1105. DOI: [https://doi.org/10.1016/S0140-6736\(66\)92194-5](https://doi.org/10.1016/S0140-6736(66)92194-5).
- [19] Vera Ruzsanyi, Jörg I. Baumbach, Stefanie Sielemann, P. Litterst, M. Westhoff, and Lutz Freitag. 2005. Detection of human metabolites using multi-capillary columns coupled to ion mobility spectrometers. *Journal of chromatography. A* 1084, 1-2, 145-151. DOI: <https://doi.org/10.1016/j.chroma.2005.01.055>.
- [20] Patrik Španěl and David Smith. 2000. Selected ion flow tube mass spectrometry analyses of stable isotopes in water: Isotopic

- composition of H_3O^+ and $\text{H}_3\text{O}^+(\text{H}_2\text{O})_3$ ions in exchange reactions with water vapor. *J Am Soc Mass Spectrom* 11, 10, 866–875. DOI: [https://doi.org/10.1016/S1044-0305\(00\)00157-4](https://doi.org/10.1016/S1044-0305(00)00157-4).
- [21] Johann Taucher, Armin Hansel, Alfons Jordan, Ray Fall, Jean H. Futrell, and Werner Lindinger. 1997. Detection of isoprene in expired air from human subjects using proton-transfer-reaction mass spectrometry. *Rapid Commun. Mass Spectrom.* 11, 11, 1230–1234. DOI: [https://doi.org/10.1002/\(SICI\)1097-0231\(199707\)11:11<1230:AID-RCM3>3.0.CO;2-Z](https://doi.org/10.1002/(SICI)1097-0231(199707)11:11<1230:AID-RCM3>3.0.CO;2-Z).
- [22] J. J. B. N. van Berkel, J. W. Dallinga, G. M. Möller, R. W. L. Godschalk, E. J. Moonen, E. F. M. Wouters, and F. J. van Schooten. 2010. A profile of volatile organic compounds in breath discriminates COPD patients from controls. *Respiratory medicine* 104, 4, 557–563. DOI: <https://doi.org/10.1016/j.rmed.2009.10.018>.
- [23] Rens van de Goor, Michel van Hooren, Anne-Marie Dingemans, Bernd Kremer, and Kenneth Kross. 2018. Training and Validating a Portable Electronic Nose for Lung Cancer Screening. *Journal of thoracic oncology : official publication of the International Association for the Study of Lung Cancer* 13, 5, 676–681. DOI: <https://doi.org/10.1016/j.jtho.2018.01.024>.
- [24] Zhenqiang Xu, Fangxia Shen, Xiaoguang Li, Yan Wu, Qi Chen, Xu Jie, and Maosheng Yao. 2012. Molecular and microscopic analysis of bacteria and viruses in exhaled breath collected using a simple impaction and condensing method. *PloS one* 7, 7, e41137. DOI: <https://doi.org/10.1371/journal.pone.0041137>.

Benutzerstudie an einer Software von Kontrastmittelinjektoren für CT & MRT

Jessica Nadine Konecki

Hochschule Reutlingen

jessica_nadine.konecki@student.reutlingen-university.de

Abstract

In Zusammenarbeit mit dem Medizinproduktehersteller *ulrich medical* wird eine User Experience und Usability Studie an der Software der im Moment eingesetzten Kontrastmittelinjektoren durchgeführt. Das Unternehmen möchte eine neue Variante eines Kontrastmittelinjektors entwickeln, der als Basis eine verbesserte Version dieser Softwares enthält. Benutzerstudien können mit den unterschiedlichsten Methoden durchgeführt werden. Das geeignete Vorgehen muss definiert und die Testpersonen in Bezug zur eingesetzten Methode ermittelt werden. Bei Medizinprodukten muss zusätzlich auf strikte Auflagen in Normen und Gesetzen geachtet werden. Die Grundlage zur Methodenauswahl bildet eine Recherche zu Usability und User Experience Vorgaben für Medizinprodukte. Die Studie wird anhand quantitativer Daten eines Usability Tests im Labor, Fragebögen zur User Experience und qualitativen Post-Test-Interviews evaluiert. In erster Linie dient diese Studie der Ermittlung von möglichen Verbesserungen, welche in der darauf folgenden Masterthesis vertieft und umgesetzt werden.

Betreuer Hochschule: Prof. Dr. Uwe Kloos
Hochschule Reutlingen
uwe.kloos@reutlingen-university.de

Betreuer Firma: Dipl.-Ing. Herbert Natterer (FH)
ulrich medical
h.natterer@ulrichmedical.com

Wissenschaftliche Vertiefungskonferenz
20. Mai 2020, Hochschule Reutlingen
Copyright 2020 Jessica Nadine Konecki

Schlüsselwörter

HCI, User Experience, Usability, summative Evaluation, Kontrastmittelinjektor, Medizinprodukte-Software, DIN9241, DIN62366, Gebrauchstauglichkeit, Benutzererlebnis

CR-Kategorien

- Human-centered computing ~ Human computer interaction (HCI) ~ HCI design and evaluation methods ~ Usability testing
- Human-centered computing ~ Human computer interaction (HCI) ~ Empirical studies in HCI

1. Einleitung

Medizinische, elektrische Geräte werden für die Überwachung und Behandlung von Patienten eingesetzt. Die Benutzungsfehler, die durch eine mangelhafte Gebrauchstauglichkeit entstehen, können fatale Folgen mit sich bringen. Werden Medizinprodukte ohne gebrauchstauglichkeitsorientierten Entwicklungsprozess entwickelt, sind sie meist nicht intuitiv verständlich und schwer zu erlernen. [2]

Auch das Benutzererlebnis, die sogenannte User Experience (UX), zählt zu einer der zu berücksichtigenden Aspekte bei der Produktentwicklung. Ist der Benutzer dem Produkt gegenüber positiv eingestellt und erfährt bei der Benutzung keine schlechten Erfahrungen, verwendet er es zum einen gerne und empfiehlt es zum anderen weiter oder kauft es sich erneut. [13]

Das Unternehmen *ulrich medical* möchte seine Injektoren weiterhin den sich stetig weiterentwickelnden Standards angleichen.

Kontrastmittelinjektoren (Abb.1 links) sind eine zuverlässige Unterstützung für die moderne bildgebende Diagnostik mit Computer (CT)- oder Magnetresonanztomographen (MRT). Sie injizieren während des Bildgebungsverfahrens Natriumchlorid-Lösungen (NaCl) und Kontrastmittel über einen Zugang am Patienten. Injektoren von *ulrich medical* können während der Bildgebung im selben Raum wie das CT oder MRT Gerät eingesetzt werden.



**Abb.1: Kontrastmittelinjektor (links);
externer Monitor (rechts)**

Mithilfe eines externen Monitors, Terminal genannt (Abb.1 rechts), der über WLAN mit dem Injektor verbunden ist, können Injektionsprogramme erstellt und ferngesteuert gestartet werden. Für diese Studie wird jeweils ein Injektor für das MRT (Max 3™) und das CT (CT motion™) untersucht.

1.1. Forschungsfragen

RQ1: Welche Kriterien gilt es bei der Bewertung der Usability und UX von Medizinprodukten zu beachten? Diese Forschungsfrage soll die Grundlagen dazu liefern, welche Eigenschaften Benutzerstudien aufweisen, wie sie durchgeführt und welche Bewertungsskalen verwendet werden.

RQ2: Welche Punkte sollten in Bezug auf UX und Usability bei den Kontrastmittelinjektoren-Softwares verbessert werden? Diese Forschungsfrage liefert Ergebnisse, die in der darauffolgenden Masterthesis bearbeitet werden.

1.2. Methodik

Für Forschungsfrage RQ1 wird mit einer Literaturrecherche gearbeitet. Dazu werden Experten im Unternehmen befragt. Um weitere geeignete Literatur zu finden,

werden Suchterme zur Usability, UX und Medizinsoftware im Online-Portal *EDDI* der Hochschulbibliothek Reutlingen gesucht.

Forschungsfrage RQ2 wird mithilfe einer empirischen Studie beantwortet. Dazu werden quantitative Daten durch einen Test im Labor und Fragebögen erhoben. Qualitative Daten werden durch die Methode *Thinking Aloud* während und semi-strukturierten Interviews nach den Benutzer-tests gewonnen.

1.3. Struktur der Arbeit

Kapitel 2 beschreibt den Begriff der Usability und welche Qualitätskriterien anhand von Normen für die Bewertung einer Medizinprodukte-Software angewendet werden können. Kapitel 3 zeigt das Selbe für die UX. In Kapitel 4 sind aus diesen Qualitätskriterien explizit Teile für die Software der Kontrastmittelinjektoren ausgewählt und zusammengefasst. Kapitel 5 zeigt die Testvorbereitung anhand der Auswahl von Testpersonen, -aufgaben und -umgebung, sowie den Testablauf. Die wichtigsten Ergebnisse der Benutzertests sind in Kapitel 6 aufbereitet. Kapitel 7 gibt eine Zusammenfassung der Studie und Kapitel 8 einen Ausblick, wie die Ergebnisse weitere Verwendung finden.

2. Usability

Usability ist ein Teilgebiet der Mensch-Maschine-Interaktion und wird am besten mit *Gebrauchstauglichkeit* oder mit *Benutzerfreundlichkeit* übersetzt. Sie beschreibt die Arbeit hin zu leicht verständlicher und schnell benutzbarer Software.[6] Die Norm DIN 9241-11 definiert, dass es das „Ziel der Gestaltung und Evaluierung gebrauchstauglicher Systeme [...] ist, die Benutzer zur effektiven, effizienten und zufriedenstellenden Erreichung ihrer Ziele [...] zu befähigen.“[6]

2.1. Qualitätskriterien von Produkten allgemein

Die Normreihe 9241 ist ein internationaler Standard, der Qualitätsrichtlinien der

Mensch-Computer-Interaktion beschreibt. [7] Die Teile 11, 110 und 112 definieren einige Qualitätskriterien für die Gestaltung interaktiver Systeme und werden in den folgenden Kapiteln näher betrachtet.

2.1.1. DIN 9241-11: Gebrauchstauglichkeit–Begriffe & Konzepte

Die Norm beschreibt, wie Usability bezogen auf die Benutzerleistung und -zufriedenheit interpretiert werden kann. Dafür können drei Qualitätskriterien genannt werden, die aus der Nutzung eines Systems resultieren: Effektivität, Effizienz und Zufriedenstellung. [6] Tabelle 1 beschreibt diese Grundsätze. Die Grundsätze gliedern sich in weitere Unterkategorien. Zum einen tragen die Grundsätze der Dialoggestaltung (DIN 9241-110) und zum anderen die Grundsätze der Informationsdarstellung (DIN 9241-112) in interaktiven Systemen dazu bei, dass

Benutzer bestimmte Ziele effektiv, effizient und zufriedenstellend erreichen können.

2.1.2. DIN 9241-110: Grundsätze der Dialoggestaltung

Die Grundsätze der Dialoggestaltung können bei der Analyse, Gestaltung und Bewertung von Dialogen angewendet werden. Als Dialog wird die Interaktion zwischen einem Benutzer und einem interaktiven System verstanden. [7] In Tabelle 2 werden sieben Qualitätskriterien für die Gestaltung und Bewertung eines Dialoges vorgestellt.

2.1.3. DIN 9241-112: Grundsätze der Informationsdarstellung

Das Anwenden der Grundsätze der Informationsdarstellung kann helfen Probleme mit der dargestellten Information zu verringern. [8] In Tabelle 3 werden sechs Grundsätze für die Gestaltung und Bewertung von dargestellter Information vorgestellt.

Tab 1. Qualitätskriterien der Usability [6]

EFFEKTIVITÄT	EFFIZIENZ
Effektivität ist die Genauigkeit und Vollständigkeit, mit denen Benutzer bestimmte Ziele erreichen. Sie bezeichnet das Ausmaß der Übereinstimmung von tatsächlichen und angestrebten Ergebnissen.	Effizienz bezeichnet die im Verhältnis zu den erreichten Ergebnissen eingesetzten Ressourcen. Diese Ressourcen umfassen z.B. Zeit, menschlicher Aufwand, Geld und Materialien.
ZUFRIEDENSTELLUNG	
Zufriedenstellung ist das Ausmaß, in dem die Reaktionen des Benutzers in Übereinstimmung mit den Benutzererfordernissen und Benutzererwartungen sind.	

Tab 2. Qualitätskriterien der Dialoggestaltung [7]

AUFGABENANGEMESSENHEIT	INDIVIDUALISIERBARKEIT
Ein System ist aufgabenangemessen, wenn es den Benutzer unterstützt, seine Arbeitsaufgabe zu erledigen.	Ein Dialog ist individualisierbar, wenn Benutzer Interaktionen und Darstellungen von Informationen an ihre Fähigkeiten und Bedürfnisse anpassen können.
ERWARTUNGSKONFORMITÄT	LERNFÖRDERLICHKEIT
Ein Dialog ist erwartungskonform, wenn er den Benutzerbelangen sowie anerkannten Konventionen entspricht.	Ein Dialog ist lernförderlich, wenn er den Benutzer beim Erlernen der Nutzung des Systems unterstützt und anleitet.

STEUERBARKEIT	FEHLERTOLERANZ
Ein Dialog ist steuerbar, wenn der Benutzer in der Lage ist, den Dialogablauf zu starten sowie seine Richtung und Geschwindigkeit zu beeinflussen.	Ein Dialog ist fehlertolerant, wenn das Arbeitsergebnis trotz erkennbar fehlerhafter Eingaben mit keinem oder geringem Korrekturaufwand erreicht werden kann.
SELBSTBESCHREIBUNGSFÄHIGKEIT	
Ein Dialog ist selbstbeschreibungsfähig, wenn für den Benutzer zu jeder Zeit offensichtlich ist, in welchem Dialog er sich befindet, welche Handlungen unternommen und wie diese ausgeführt werden können.	

Tab 3. Qualitätskriterien der Informationsdarstellung [8]

ENTDECKBARKEIT	KOMPAKTHEIT
Dargestellte Informationen sind entdeckbar, wenn die Informationen als vorhanden wahrgenommen werden.	Die Informationsdarstellung ist kompakt, wenn nur die notwendigen Informationen dargestellt werden.
INTERPRETIERBARKEIT	ABLENKUNGSFREIHEIT
Informationen sind eindeutig, wenn sie so verstanden werden, wie es vorgesehen ist.	Informationen sind ablenkungsfrei, wenn erforderliche Informationen wahrgenommen werden, ohne dass weitere dargestellte Informationen stören.
UNTERSCHIEDBARKEIT	KONSISTENZ
Informationen sind unterscheidbar, wenn eigenständige Elemente oder Gruppen genau voneinander unterscheidbar sind.	Dargestellte Informationen sind konsistent, wenn Informationselemente mit ähnlichem Zweck ähnlich dargestellt werden.

2.2. Qualitätskriterien von Medizinprodukten

Neben den bereits erwähnten Qualitätskriterien Effizienz, Effektivität und Zufriedenstellung, ist die Sicherheit, also die Vermeidung von Benutzerfehlern eines der wichtigsten Ziele von Usability in der Medizintechnik. Usability wird hier im Zusammenhang mit der Sicherheit des Benutzers, Patienten und Dritten verstanden und kann mit *sicherer Gebrauch* übersetzt werden.

2.2.1. Definition Medizinprodukt

Die Medizinprodukte-Verordnung (MDR) definiert ein Medizinprodukt als „[...] ein Gerät, [...] das dem Hersteller zufolge für Menschen bestimmt ist und [...] einen oder mehrere [...] medizinische[n] Zwecke erfüllen soll und dessen bestimmungsgemäße Hauptwirkung im oder am menschlichen Körper weder durch pharmakologische¹ oder immunologische² Mittel noch metabolische³ erreicht wird, dessen Wirkungsweise aber durch solche Mittel unterstützt werden kann.“[1]

¹ die Wirkung von Medikamenten betreffend

² das Immunsystem betreffend

³ im Stoffwechsel entstanden

2.2.2. Usability als Synonym für Sicherheit

Im Folgenden werden die vermeintlich wichtigsten Regularien vorgestellt, deren Grundsätze beschrieben und beurteilt, in wie weit diese einen Beitrag zur Anwendung von Usability auf Medizinprodukte leisten.

Die MDR legt strenge (sicherheitsrelevante) Auflagen für Medizinproduktehersteller fest. [1] Ein eigenes Kapitel über Usability ist jedoch nicht vorhanden. Es findet sich lediglich ein Verweis auf die Normen DIN 6060-1 und DIN 62366.

Die DIN 62304 beschreibt den Lebenszyklus-Prozess von Medizinprodukte-Software. Über die Bewertung von Usability einer Medizinsoftware werden in dieser Norm keine expliziten Angaben gemacht.[3]

Die Norm DIN 60601-1-6 beschreibt einen auf die Bedürfnisse von Herstellern medizinischer Produkte zugeschnittenen gebrauchstauglichkeitsorientierten Entwicklungsprozess.[2]

DIN 62366-1 beschreibt, wie auch die DIN 60601-1-6, nur die Aktivitäten, die ein gebrauchstauglichkeitsorientierter Entwicklungsprozess enthalten muss.[4] Sie gibt weder einen konkreten Prozess vor, noch nennt sie Methoden, die angewendet werden können oder müssen. Dafür ist der technische Report IEC TR 62366-2 zuständig. Der technische Report ist ein Leitfaden zur Anwendung des Usability Engineering auf Medizinprodukte. Er bietet den Herstellern eine Richtlinie, wie sie die Prinzipien der Usability in ihre eigenen Entwicklungsprozesse integrieren können.[5] Er kann als Anleitung zur Anwendung der DIN 62366-1 gesehen werden und beschreibt wie sich Usability auf Attribute wie Aufgabeneffizienz und Zufriedenheit der Benutzer auswirkt. Es werden aber nur Anforderungsbeispiele genannt und auf die sieben Grundsätze der Dialoggestaltung der Normreihe 9241 verwiesen.

Wenn in den aufgeführten Verordnungen und Normen über den Umgang mit Produkten

gesprochen wird, geht es vor allem um die Sicherheit. Der Zweck ist hauptsächlich einen Entwicklungsprozess zur Verfügung zu stellen, der sichere Medizinprodukte-Software erzeugt. Welche Qualitätskriterien dafür berücksichtigt oder gemessen werden sollen, wird aber nicht vorgegeben.

2.2.3. DIN 62366 in Bezug zur DIN 9241

Die DIN 62366 trägt von allen untersuchten medizinischen Regularien am konkretesten zur Anwendung der Usability auf Medizinprodukte bei. Während die DIN 62366 aber keine Anforderung an die Gestaltung von Benutzerschnittstellen stellt, gibt die DIN 9241 ganz direkte Gestaltungsrichtlinien vor. Ein großer Kritikpunkt an der DIN 62366 ist, dass die Norm zwar auf die DIN 9241 verweist aber keine konkreten Forderungen daraus übernimmt. Die Qualitätskriterien der DIN 9241 lassen sich größtenteils auf alle interaktiven Systeme und damit auch auf Medizinprodukte übertragen. Allerdings berücksichtigt die DIN 9241 im Gegensatz zur DIN 62366 nicht die Sicherheit von Benutzern und Patienten.

3. User Experience

Die UX, auch *Nutzungserlebnis* genannt, erweitert den Begriff der Usability um ästhetische und emotionale Faktoren. UX ist die Wahrnehmung und Reaktion einer Person vor, während und nach der Benutzung eines Systems. Sie schließt alle Emotionen, Vorlieben, Wohlbefinden und Verhaltensweisen mit ein.[6] Was der Benutzer vor der Benutzung erwartet, wie es sich anfühlt das Produkt zu benutzen und die Wahrnehmung nach der Benutzung sind Teil der UX.

3.1. Qualitätskriterien von Produkten allgemein

Die Norm 9241-210 verwendet ein breit ausgelegtes Konzept der Usability und unterstellt sie der UX.[9] Die Norm merkt aber auch an, dass es für sicherheitskritische Systeme wichtiger sein könnte, die Usability betreffende Effektivität und Effizienz des Systems sicherzustellen als die Benutzer-

präferenzen. Die Norm gibt keine konkreten Anforderungen, wie die UX umgesetzt und bewertet werden kann, sondern besagt, dass während der Gestaltung eines Systems die gesamte UX berücksichtigt werden muss und verweist dabei auf die in Abschnitt 2.1 behandelten Grundsätze der DIN 9241. Wenn die Grundsätze der Usability erfüllt sind, wird demnach auch die UX angemessen ausfallen.

3.2. Funktionale, hedonische und produktspezifische Qualitätskriterien

Nach Schrepp [13] können für die Bewertung von Produkten Qualitätskriterien in drei Kategorien eingeteilt werden: funktionale Grundsätze der Usability, hedonische und

produktspezifische Grundsätze der UX, die den Usability Begriff erweitern.

Funktionale Qualitätskriterien zeigen, wie gut bzw. schlecht Benutzer ihre Aufgaben mit einem Produkt bearbeiten können. Vor allem sind das die in Kapitel 2.1 beschriebenen Grundsätze der Usability.

Hedonische Qualitätskriterien sind Merkmale, die sich nicht unbedingt an der Bearbeitung von Aufgaben orientieren und sind in Tabelle 4 aufgeführt.

Produktspezifische Qualitätskriterien sind Merkmale, die für jedes Produkt verschieden und vom Nutzungskontext und Benutzer abhängig sind. Beispiele der gängigsten Kriterien sind in Tabelle 5 aufgeführt.

Tab 4. Beschreibung der hedonischen Qualitätskriterien [13]

ORIGINALITÄT / INNOVATION	SCHÖNHEIT
Ziel einer originellen Gestaltung ist es, Aufmerksamkeit auf sich zu ziehen, z.B. durch neuartige Funktionen oder originelle visuelle Gestaltung.	Die Schönheit eines Produkts kann einen positiven Einfluss auf die wahrgenommene Usability haben, obwohl das Produkt noch nicht benutzt wurde.
STIMULATION (FUN-OF-USE)	WERTIGKEIT
Stimulation soll dem Benutzer Spaß oder Freude und keine Langeweile bei der Benutzung eines Produkts bringen.	Der Besitz eines Produkts hat für den Benutzer über die reine Nutzung des Produkts hinausgehende Qualität.
VALENZ (ATTRAKTIVITÄT)	VERBUNDENHEIT
Allgemeine Zufriedenheit bzw. Unzufriedenheit mit einem Produkt.	Verbundenheit sichert dem Hersteller immer wiederkehrende Käufe oder dauerhafte Nutzung eines Systems.

Tab 5. Beschreibung der produktspezifischen Qualitätskriterien [13]

AKUSTIK	HAPTİK
Wahrnehmung, wie sich ein Produkt anhört.	Wahrnehmung, wie sich ein Produkt anfühlt
INHALTSQUALITÄT	VERTRAUEN/SICHERHEIT
Darstellung von wichtigen, interessanten oder aktuellen Informationen.	Vertrauen ist bei Systemen, die persönliche und/oder sicherheitsrelevante Informationen verarbeiten ein wichtiges Qualitätskriterium.

4. Anforderungen für die Bewertung der Software

Aufgrund der oben erörterten Ergebnisse, dass medizinische Normen für konkrete Anweisungen zur Erreichung einer guten UX und Usability auf die Grundsätze der Normreihe DIN 9241 verweisen, werden für die Bewertung der Injektoren-Software diese Qualitätskriterien verwendet. Ergänzend außerdem eine Auswahl der hedonischen und produktspezifischen Qualitätskriterien.

4.1. Bewertung der Usability

Die Usability wird anhand der Qualitätskriterien aus Kapitel 2.1 bewertet. Mithilfe passender Metriken in Form eines Fragebogens müssen die Testpersonen nach dem Nutzertest die funktionalen Qualitätskriterien bewerten. Abbildung 2 zeigt ein Beispiel für eine Metrik zur Bewertung der Ablenkungsfreiheit. Der Usability-Fragebogen beinhaltet 42 solcher Bewertungsmetriken und ist speziell für die Injektoren-Software entworfen. Der Fragebogen ist so aufgebaut, dass bei jeder Metrik die negative Bewertungsmöglichkeit links und die positive rechts der fünf Kreise steht. Eine Metrik kann mit einer Punktzahl von eins bis fünf bewertet werden. Je höher also die Punktzahl, umso besser schneidet ein Qualitätskriterium ab. Insgesamt können pro Testperson 210 Punkte erreicht werden.

4.2. Bewertung der UX

Die UX wird zusätzlich zur Usability anhand von hedonischen und produktspezifischen Qualitätskriterien bewertet. Da diese immer produktabhängig sind, wird eine Auswahl aus Kapitel 3.3 verwendet: Originalität/Innovation, Zufriedenheit, Stimulation (Fun-

of-Use), Wertigkeit, Vertrauen/Sicherheit, Schönheit und Inhaltsqualität.

Mithilfe passender Metriken in Form eines Fragebogens müssen die Testpersonen nach dem Nutzertest die UX bewerten. Der UX Fragebogen beinhaltet 20 solcher Bewertungsmetriken und ist speziell für die Injektoren-Software entworfen. Die Bewertung erfolgt nach dem gleichen Prinzip, wie im vorherigen Abschnitt zum Usability-Fragebogen beschrieben. Insgesamt können pro Testperson 100 Punkte erreicht werden.

5. Testphase

Dieses Kapitel zeigt das Vorgehen während der Testplanung und -durchführung. Eine detaillierte Testbeschreibung ist wichtig. Zu oberflächlich formulierte Testpläne können am Ende zu Problemen bei der Rückverfolgbarkeit und Auswertung führen. [11]

5.1. Benutzergruppen

- Laien: Testpersonen mit geringem bis keinem nutzungskontext-spezifischen⁴ Hintergrund.

Die Gruppe kann den besten Beweis liefern, dass die neue Software (Ergebnis der an diese Studie anschließenden Masterthesis) intuitiv und sogar für Benutzer ohne nutzungskontext-spezifischen Hintergrund bedienbar ist.

- Klinikpersonal: Testpersonen mit einem mittleren bis hohen nutzungskontext-spezifischen Hintergrund

Diese Gruppe repräsentiert die Zielgruppe, also die regulären, täglichen Benutzer des Produkts in Kliniken und radiologischen Abteilungen.



Abb.2: Beispiel einer Bewertungsmetrik aus dem Usability-Fragebogen

⁴ *Nutzungskontext-spezifisch* bedeutet, dass Testpersonen den Nutzungskontext, für den das Produkt entwickelt wurde, teilweise bis vollständig erfüllen. Das heißt, dass sie eine medizinische Ausbildung im Bereich der medizinisch-technischen Assistenz (MT(R)A) oder ähnlichem aufweisen.

- Experten: Testpersonen, mit einem professionellen Nutzungsgrad (Service- und Schulungspersonal).

Diese Gruppe ist ausgewählt, um durch die Testergebnisse einen Vergleichswert für die beiden anderen Gruppen zu erhalten.

5.1.1. Anzahl der Testpersonen

Lewis berechnet in seiner Arbeit mit einer Formel die Anzahl der Testpersonen anhand der Fehler, die mit dem Test gefunden werden sollen und der Wahrscheinlichkeit, dass ein Benutzer auf einen Fehler trifft. [12,10] Für die Injektor-Software ist jedoch nicht bekannt, wie hoch die Wahrscheinlichkeit ist. Sauro rät in diesem Fall dazu einen Prozentsatz zu schätzen.[12] Aufgrund von früheren, vom Unternehmen vorgenommenen Usability Tests wird geschätzt, dass die Wahrscheinlichkeit auf einen Fehler zu treffen bei 30% liegt. Die Tests sollen 95% der Fehler aufdecken, weshalb die berechnete Mindestanzahl an Testpersonen 8,4 beträgt. Es wird auf zehn aufgerundet, um einerseits die späteren Berechnungen besser durchführen zu können und andererseits die Anzahl auf die zuvor definierten drei Benutzergruppen besser aufteilen zu können: vier Laien, drei Personen für das Klinikpersonal und drei Experten.

5.2. Testaufgaben

Für den Test werden für die Injektorsysteme Max 3TM und CT motionTM jeweils Testaufgaben definiert. Die Aufgaben sind in Kategorien eingeteilt: *Systemeinstellungen*, *Injektionsprogrammliste* und *-erstellung/-änderung*. Aufgaben zu den Systemeinstellungen zeigen, wie sich die Testpersonen im Einstellungs-Menü zurecht finden. Aufgaben zur Injektionsprogrammliste geben Aufschluss, wie die Testpersonen mit der Verwaltung der Programme umgehen. Aufgaben zur Erstellung von Programmen zeigen, wie die Testpersonen zum einen vorgegebene Daten in ein Programm übertragen und wie sie Informations- und Warnungsmeldungen beheben.

5.3. Testablauf

Die Studie basiert auf einem Test im Labor mit Fragebogen und Post-Test-Interview. Die Nutzertests finden im Testfeld des Unternehmens statt. Jede Test-Session dauert ca. 45min und wird für die spätere Auswertung gefilmt. Die Testperson bearbeitet die vorliegenden Aufgaben nur mithilfe des Terminals. Anschließend füllt sie die Fragebögen aus. Der Moderator gibt keine Hinweise und greift nur in Ausnahmefällen in den Test ein.

6. Ergebnisse

Die Ergebnisse werden in fünf Teilen ausgewertet:

- Auswertung der Usability Fragebögen
- Auswertung der Effizienz und Effektivität mithilfe der Videoaufnahmen
- Auswertung der UX Fragebögen
- Vergleich der Benutzergruppen
- Auswertung der Post-Test-Interviews

Innerhalb der Bewertungen sind die Werte weitestgehend homogen, weshalb auf die Angabe von Streuungswerten zugunsten der Übersichtlichkeit verzichtet wird.

6.1. Usability

Die Usability des Max 3TM erreicht 69% und die des CT motionTM 76% der erreichbaren Punkte. Bei der Bewertung der Qualitätskriterien erhält der CT motionTM in den meisten Bereichen eine etwas bessere Punktzahl, wie in Abbildung 3 zu sehen ist. Die Unterschiede zwischen den beiden Injektoren liegen jedoch unterhalb von 0,55 Punkten.

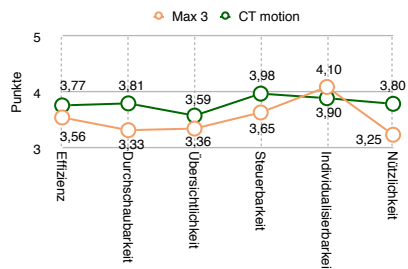


Abb.3: Ø Bewertung der Usability-Qualitätskriterien

Die Auswertung der am niedrigsten bewerteten Metriken des Max 3TM zeigt, dass die Software schwer zu bedienen ist (2,7), die Testpersonen sich unsicher fühlen (2,6), Informationen erst nach langem Suchen (2,6) und sehr langsam gefunden werden (2,7). Die Auswertung der schlechtesten Metriken des CT motionTM zeigt, dass die Software schwer zu bedienen ist (3,4), die Testpersonen sich nicht zurechtfinden (3,4), Informationen sehr langsam finden (3,4) und nicht genau wissen, was als nächstes zu tun ist (3,4).

6.2. Effektivität

Die Effektivität der Software wird mithilfe der Bewertungsmetriken der Vollständigkeit und Genauigkeit gemessen. Die Genauigkeit richtet sich danach, inwieweit die Lösung der Aufgabe korrekt ist. Dabei wird für diese Studie zwischen drei Übereinstimmungsgraden unterschieden, welche bei der Auswertung unterschiedlich gewichtet werden: nicht, teilweise und vollständig korrekt.

Tab 6. Genauigkeit der Aufgaben

	Max3 TM	CT motion TM
Laien	86 %	95 %
Klinikpersonal	90 %	95 %
Experten	90 %	98 %

Tabelle 6 zeigt, wie hoch die Genauigkeit der gelösten Aufgaben jeder Benutzergruppe ist.

Die Vollständigkeit definiert, ob die Testpersonen in der Lage sind, alle vorgegebenen Aufgaben vollständig korrekt zu lösen. Sie ist damit ein Teilaspekt der Genauigkeit.

Tab 7. Vollständigkeit der Aufgaben

	Max3 TM	CT motion TM
Laien	80%	91%
Klinikpersonal	87%	91%
Experten	80%	97%

Tabelle 7 zeigt, wie hoch der Anteil an den vollständig korrekt gelösten Aufgaben jeder Benutzergruppe ist.

Der Anteil der Vollständigkeit liegt bei allen Benutzergruppen unterhalb des Anteils der Genauigkeit. Das bedeutet, dass Aufgaben nicht immer vollständig korrekt, sondern auch nur teilweise oder nicht korrekt gelöst werden. Wie in Abbildung 4 zu sehen ist, schneidet die Software des CT motionTM bei beiden Metriken besser ab.

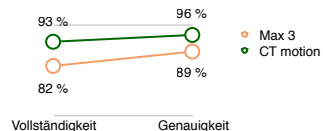


Abb.4: Vergleich der Vollständigkeit und Genauigkeit

6.3. Effizienz

Die Effizienz wird zusätzlich zum Fragebogen über die Dauer und Anzahl an Klicks gemessen, die eine Testperson benötigt, um die Aufgaben zu lösen. Für jede Aufgabe sind die Idealwerte der maximalen Anzahl an Klicks und der Höchstdauer ermittelt worden, die ohne Fehler zum Lösen der Aufgabe benötigt werden.

Tabelle 8 zeigt die Idealwerte der Dauer und Klicks zum Lösen aller Aufgaben im Vergleich zum erreichten Durchschnittswert der Testpersonen.

Tab 8. Vergleich der Effizienz

	Idealwert	Ø Testpersonen
Max 3TM	1,85 Min 91 Klicks	15,2 Min 288 Klicks
CT motionTM	2,25 Min 118 Klicks	11,5 Min 255 Klicks

Abbildung 5 zeigt die durchschnittliche Dauer und Abbildung 6 die Anzahl an Klicks einer Testperson für die drei Aufgabenkategorien im Vergleich zum Idealwert. In den Abbildungen ist zu sehen, dass beide Softwares bei der Dauer, sowie der Anzahl

an Klicks deutlich über dem Idealwert liegen. Die Anzahl an unnötigen Klicks ist demnach sehr hoch.

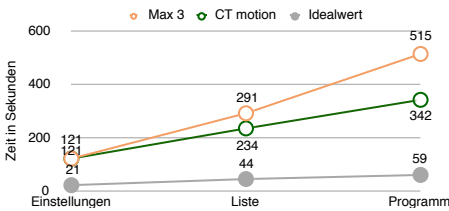


Abb.5: Ø Dauer einer Testperson für Aufgabenkategorien

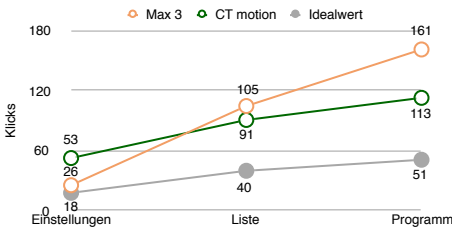


Abb.6: Ø Anzahl an Klicks einer Testperson für Aufgabenkategorien

6.4. UX

Die UX des Max 3™ erreicht 75% und die des CT motion™ 78% der erreichbaren Punkte. Im Vergleich erreicht der CT motion™ bei der Hälfte der Qualitätskriterien eine bessere Bewertung als der Max 3™, wie in Abbildung 7 zu sehen ist. Die Unterschiede zwischen den beiden Injektoren liegen in allen Bereichen jedoch unterhalb von 0,45 Punkten.

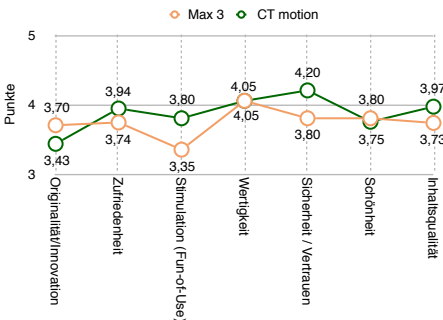


Abb.7: Ø Bewertung der UX-Qualitätskriterien

Die Auswertung der am niedrigsten bewerteten Metriken des Max 3™ zeigt, dass die Software schwer zu bedienen (3,3) ist, die Testpersonen unsicher (3,2), frustriert (2,9) und unzufrieden mit dem Ergebnis ihrer bearbeiteten Aufgaben (3,2) sind. Die Auswertung der schlechtesten Bewertungsmetriken des CT motion™ zeigt, dass die Software schwer zu bedienen (3,6) und das Design gewöhnlich (3,3), phantasielos (3,4) und langweilig (3,6) ist.

6.5. Benutzergruppenvergleich

Obwohl die Benutzergruppen unterschiedliche Ausbildungen und Vorkenntnisse haben, fallen vor allem die Bewertungen der Laien und Experten sehr ähnlich aus.

Abbildung 8 zeigt für jede Benutzergruppe die durchschnittliche Bewertung der Usability und Abbildung 9 der UX. Die Benutzergruppe der Laien und der Experten bewerten den Max 3™ in beiden Bereichen mit fast der selben Punktzahl.

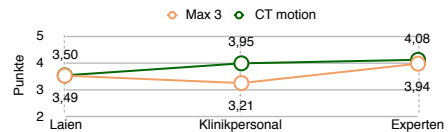


Abb.8: Ø Bewertung der Usability

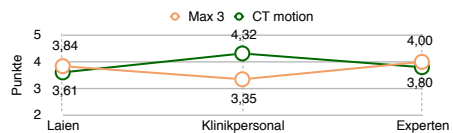


Abb.9: Ø Bewertung der UX

Die Usability des Max 3™ wird von allen Benutzergruppen schlechter bewertet als die des CT motion™, wie Abbildung 8 zeigt. Die Laien und Experten bewerten die User Experience des Max 3™ aber besser als das Klinikpersonal, wie in Abbildung 9 zu sehen.

Die Unterschiede zwischen Max 3™ und CT motion™ liegen in den Diagrammen bei den Laien und Experten jeweils unter 0,14 (Usability) und 0,23 (UX) Punkten. Der Bewertungsunterschied des Klinikpersonals liegt hingegen bei 0,74 (Usability) und 0,97 (UX) Punkten.

6.6. Post-Test-Interview

Der größte Kritikpunkt an der Software des Max 3TM ist, dass es zu wenig textliche Beschriftungen von Buttons und Menüs gibt. Es ist schwer sich zurecht zu finden. Symbole auf Buttons sind meistens nicht eindeutig. Das zweite Problem sind Informationen und Warnungsmeldungen, die nicht sofort erkennbar und eindeutig interpretierbar sind. An der Software des CT motionTM stört die Testpersonen, dass es zu viele Schritte und unnötige Eingaben gibt. Es werden Informationen angezeigt, die man nicht benötigt. Außerdem sind auch hier die Symbole unverständlich.

Sechs Testpersonen würden die Software des Max 3TM der Software des CT motionTM vorziehen, wie Tabelle 9 zeigt.

Tab 9. Entscheidung für eine Software

	Max3 TM	CT motion TM
Laien	3	1
Klinikpersonal	1	2
Experten	2	1

7. Zusammenfassung

Wie genau Usability und UX in der Medizintechnik umgesetzt und Gestaltungslösungen aussehen können, gibt nur die nicht speziell auf Medizinprodukte ausgelegte DIN 9241 vor. Für Medizinprodukte gelten also die gleichen Qualitätskriterien, wie für andere Produkte auch. Allerdings muss die Sicherheit von Benutzern und Patienten strenger beachtet werden. Damit ist Forschungsfrage RQ1 beantwortet.

Die folgenden Erkenntnisse beantworten Forschungsfrage RQ2. Die Ergebnisse der Benutzerstudie haben ergeben, dass der Max 3TM nahezu in allen Qualitätskriterien schlechter bewertet wird. Es ist aber zu berücksichtigen, dass die Unterschiede meistens sehr gering sind. Die Usability des Max 3TM hat Defizite bei der Durchschaubarkeit und Übersichtlichkeit. Das bedeutet, dass

dargestellte Informationen nicht selbsterklärend und benötigte Informationen schwer zu finden sind. Die UX des Max 3TM schneidet bei der Stimulation, also der Freude am Bedienen, am schlechtesten ab. Bei der Usability des CT motionTM wird auch die Übersichtlichkeit bemängelt. Das heißt, dass benötigte Informationen schwer zu finden sind und dargestellte Informationen nicht verstanden werden. Die UX des CT motionTM wird in den visuellen Bereichen Originalität/Innovation und Schönheit am niedrigsten bewertet.

Die Effizienz der Software liegt für beide Kontrastmittelinjektoren bei der Dauer und Anzahl an Klicks weit über dem Idealwert. Die meisten Schwierigkeiten haben die Testpersonen beim Erstellen eines Programms und Beheben von Informations- und Warnungsmeldungen. Die Effektivität der Software des CT motionTM schneidet besser ab als die des Max 3TM, da die eingesetzte Benutzerführung eine nicht korrekte Aufgabenlösung fast nicht ermöglicht. Die Genauigkeit und Vollständigkeit ist bei beiden Softwares mit über 82% sehr gut und damit sind sie effektiv. Die Dauer und Anzahl an Klicks liegen aber durchschnittlich 460% über dem Idealwert, was die Softwares nicht sehr effizient macht.

Obwohl nach der quantitativen Auswertung die meisten Daten des Max 3TM minimal schlechter ausfallen, entscheiden sich dennoch 60% der Testpersonen für diese Software. Nachdem die Bewertung der Usability unter den Benutzergruppen (außer dem Klinikpersonal) sehr ähnlich ausfällt, kann diese Entscheidung nur auf die Ergebnisse der UX zurückzuführen sein. Das Design der Benutzeroberfläche des CT motionTM wird hier als gewöhnlich, phantasielos und langweilig beschrieben. Demnach muss das Erscheinungsbild der Benutzeroberfläche einen größeren Einfluss auf die Entscheidung der Testpersonen haben als die Bedienbarkeit der Software und das Erfüllen von Aufgaben.

Die präsentierten Ergebnisse sind eine Zusammenfassung der wichtigsten Qualitäts-

kriterien. In einer ausführlichen Dokumentation gehen die Ergebnisse weiter in die Tiefe.

8. Ausblick

Die erarbeiteten Ergebnisse werden in der auf diese Studie folgenden Masterthesis weiter konkretisiert. Die Ursachen der von den Testpersonen kritisierten Punkte müssen direkt in der Softwareoberfläche gefunden werden. Mit diesen Anforderungen werden Verbesserungen für ein neues Gestaltungskonzept erarbeitet, welches mithilfe eines Prototypen getestet und dem Unternehmen *ulrich medical* vorgestellt wird. Ziel ist es, die Software so selbsterklärend zu entwerfen, dass sie vor allem für das Klinikpersonal effizient, effektiv und zufriedenstellend benutzbar ist, aber auch für Laien gut bedienbar sein kann.

9. Literaturverzeichnis

- [1] Backhaus, Claus; Benad, Nadine; Lau, Hans-Joachim; Pleiss, Thomas: MDR & Co - Eine Vorschriftensammlung zum europäischen Medizinprodukterecht, Köln: TÜV Media GmbH, 2017
- [2] DIN e.V.: DIN EN 60601-1-6:2016-02 (VDE 0750-1-6), Medizinische elektrische Geräte – Teil 1-6: Allgemeine Festlegung für die Sicherheit einschließlich der wesentlichen Leistungsmerkmale – Ergänzungsnorm: Gebrauchstauglichkeit (IEC 60601-1-6:2010 + A1:2013), Berlin: Beuth Verlag
- [3] DIN e.V.: DIN EN 62304 (VDE 0750-101), Medizingeräte-Software – Software-Lebenszyklus-Prozesse (IEC/62A/838/CD:2012), Berlin: Beuth Verlag
- [4] DIN e.V.: DIN EN 62366-1:2017-07 (VDE 0750-241-1), Medizinprodukte – Teil 1: Anwendung der Gebrauchstauglichkeit auf Medizinprodukte (IEC 62366-1:2015 + COR1:2016), Berlin: Beuth Verlag
- [5] DIN e.V.: IEC TR 62366-2:2016-04, Medical Devices – Part 2: Guidance on the application of usability engineering to medical devices, Berlin: Beuth Verlag
- [6] DIN e.V.: DIN EN 9241-11:2018-11, Ergonomie der Mensch-System-Interaktion – Teil 11: Gebrauchstauglichkeit: Begriffe und Konzepte, (ISO 9241-11:2018), Berlin: Beuth Verlag
- [7] DIN e.V.: DIN EN 9241-110:2008-09, Ergonomie der Mensch-System-Interaktion – Teil 110: Grundsätze der Dialoggestaltung (ISO 9241-110:2006), Berlin: Beuth Verlag
- [8] DIN e.V.: DIN EN 9241-112:2017-08, Ergonomie der Mensch-System-Interaktion – Teil 112: Grundsätze der Informationsdarstellung (ISO 9241-112:2017), Berlin: Beuth Verlag
- [9] DIN e.V.: DIN EN ISO 9241-210:2011-01, Ergonomie der Mensch-System-Interaktion – Teil 210: Prozess zur Gestaltung gebrauchstauglicher interaktiver Systeme (ISO 9241-210:2010), Berlin: Beuth Verlag
- [10] Lewis, James R.: Sample sizes for usability tests: mostly math, not magic. Interactions 13, 29-33, USA, New York: ACM, 2006
- [11] Rubin, Jeffrey; Chisnell Dana: Handbook of Usability Testing – How to Plan, Design and Conduct Effective Tests, 2. Ausgabe, USA, Indianapolis: Wiley Publishing Inc., 2008
- [12] Sauro, Jeff: A Practical Guide to Measuring Usability – Quantifying the Usability of Websites & Software, USA, Denver: Measuring Usability LLC Publication, 2010
- [13] Schrepp, Martin: User Experience mit Fragebögen messen, USA, North Charleston: CreateSpace Verlag 2018

Regulatorische Anforderungen an eine Qualitätsmanagement-Software für Medizinprodukte

Laura Discher

Hochschule Reutlingen

Laura.Discher@Student.Reutlingen-University.de

Abstract

Medizinprodukte sind Gegenstände, Stoffe oder Software mit medizinischer Zweckbestimmung für die Anwendung am Menschen. Diese werden von Medizinprodukteherstellern entwickelt und auf den Markt eingeführt. Da die falsche Anwendung von Medizinprodukten bei Menschen zu Verletzbarkeit des menschlichen Körpers führen kann, ist eine angemessene Qualität der Medizinprodukte zu gewährleisten. Um die Sicherstellung der Qualität einzuhalten sind Medizinproduktehersteller verpflichtet sich an die Medizinprodukteverordnung (MDR) zu halten. Für risikoreiche Produkte ist ergänzend die Nutzung eines Qualitätsmanagementsystems (QMS) verpflichtend. Dieses steuert die Struktur, Verantwortlichkeiten, Verfahren und Prozesse des Unternehmens, die für die Medizinprodukteentwicklung notwendig sind. In Zeiten der Digitalisierung werden Softwarelösungen eingesetzt, um die zeitaufwendigen Dokumentations- und Administrationstätigkeiten im QMS zu reduzieren und die Prozesse zu optimieren. Mit der Einführung einer Software wird ein QMS in der Pra-

xis auch als elektronisches QMS (eQMS) bezeichnet. Weiterhin muss das gesamte QMS mit den Regularien konform sein. Deshalb ist das Ziel dieser Arbeit mithilfe den regulatorischen Anforderungen herauszuarbeiten, welche Vorgaben bei der Einführung eines eQMS zu beachten sind und wie diese erfüllt werden können. Diese Arbeit bezieht sich auf die regulatorischen Vorgaben aus der MDR und der ISO 13485. Die Norm beinhaltet Anforderungen an ein QMS von Medizinprodukten.

Schlüsselwörter

Qualitätsmanagementsystem, Medizinprodukte, Qualitätsmanagement-Software, regulatorische Anforderungen, DIN EN ISO 13485, Medizinprodukteverordnung

CR-Kategorien

- Computer Applications ~ Life and medical sciences
- Software Engineering ~ Requirements/Specifications

1 Einleitung

Die Medizintechnik ist in Zeiten der Digitalisierung, Big Data und Automatisierung ein stets wachsender Teil des deutschen Gesundheitswesens und dessen Wirtschaft. Dabei richten sich die Entwicklungen, Dienstleistungen und Produkte am demografischen Wandel im Hinblick auf die steigende Patientenorientierung und Personalisierung aus. Neben den Chancen dieser Branche wie der Innovationswelle und der zunehmenden Nachfrage müssen Hersteller mit den damit verbundenen Herausforderungen umgehen

Betreuer Hochschule: Prof. Dr. Kücherer
Hochschule Reutlingen
Christian.Kuecherer@Reutlingen-University.de

Wissenschaftliche Vertiefungskonferenz
20. Mai 2020, Hochschule Reutlingen
Copyright 2020 Laura Discher

können. Der Umfang an europäischen Regularien wie den Anforderungen aus der Medizinprodukteverordnung (Medical Device Regulation, MDR) [1], und den nationalen Gesetzen steigt mit dem Einsatz von Softwarelösungen. Als nationales Gesetz ist das Medizinprodukte-Durchführungsgesetz (MPDG) zu nennen [6].

1.1 Motivation

Für die Erfüllung der gesetzlichen Vorgaben setzen Medizinproduktehersteller Qualitätsmanagementabteilungen ein, die mithilfe eines Qualitätsmanagementsystems (QMS) die Grundlage für einen organisierten Produktlebenszyklus bilden. Um den durch die Regularien entstehenden Zeitaufwand an Dokumentations- und Administrationstätigkeiten in der Medizintechnik-Branche zu reduzieren, führen Hersteller elektronische QMS (eQMS) in Form einer funktionsübergreifenden Software ein. Damit findet eine Umstellung von manuellen Prozessabläufen mit papierbasierter Dokumentation und handschriftlicher Unterschrift zu automatisierten Prozessen mit elektronischer Signatur statt. Im Zuge dessen kann der Fokus gezielter auf die wesentlichen Aufgaben gelenkt werden.

Die Vorgaben der Gesetze beziehen sich neben den Medizinprodukten auch auf QMS. Mit einer Zertifizierung nach dem aktuellen Stand der Technik gemäß DIN EN ISO 13485 [2] können Medizinproduktehersteller die Anforderungen an ein QMS gewährleisten. Dies stellt die Grundlage an die Konformität der Medizinprodukte, dessen Sicherheit und Qualität dar. Nach DIN EN ISO 9000 [3] ist mit dem Begriff Qualität die Übereinstimmung der Produkte und Dienstleistungen mit deren vorgesehenen Zweckbestimmung und Leistung zu verstehen. Der Nutzen und Erwartungen der Anwender steht dabei im Vordergrund [3].

Für Medizinproduktehersteller ist es deshalb notwendig Softwarelösungen so einzuführen, dass ein QMS weiterhin der Konformität der Gesetze entspricht und damit Medizinprodukte qualitäts- und rechtskonform herge-

stellt werden. Dafür müssen die Anforderungen aus der Breite an Regularien zunächst allgemein herausgearbeitet werden. Im Anschluss erfolgt eine konkrete Betrachtung der Anforderungen im Rahmen einer Softwareeinführung für ein QMS. Mit Anforderung ist im regulatorischen Kontext ein verpflichtendes Erfordernis zu verstehen. Dies kann für die Produktentwicklung bedeuten, diese nach optimaler Zweckbestimmung zu produzieren [3].

1.2 Zielsetzung

Ziel dieser Arbeit ist es die Notwendigkeit und der grundlegende Aufbau eines QMS in der Medizintechnik-Branche zu definieren. Mit dieser Grundlage wird die Basis für die Einführung und Nutzung eines eQMS erzielt. Dafür werden in diesem Artikel die Anforderungen der DIN EN ISO 13485 [2] und der MDR [1] an ein QMS systematisch herausgearbeitet.

Mithilfe des *Technischen Berichts* ISO 80002-2 [4] und Ansätzen von wissenschaftlichen Arbeiten wird untersucht, wie die normgerechte Nutzung einer Software im QMS einzuführen und zu validieren ist. Diese Kenntnisse und Ergebnisse können zukünftig für eQMS-Projekte in der Praxis hilfreich sein. Daraus ergibt sich die zentrale Forschungsfrage welche Änderungen und Anforderungen bei der Einführung einer Softwarelösung im Rahmen des QMS eines Medizinprodukteherstellers zu erfüllen sind.

2 Medizinprodukte

Um den Bedarf und die Verpflichtung der Einrichtung eines QMS zu verdeutlichen, werden in diesem Kapitel die Grundlagen der Medizinprodukte erläutert.

2.1 Definition und Funktion

Als Medizinprodukte werden Gegenstände, Produkte oder Software bezeichnet, die für die therapeutische und diagnostische medizinische Zweckbestimmung für Menschen zum Einsatz kommen [10]. Desweiteren gehören nach § 3 des Medizinproduktegesetzes alle miteinander verknüpften oder einzelne

Instrumente, Apparate sowie Stoffe und Zubereitungen aus Stoffen zu den Medizinprodukten. Als Beispiel sind Implantate, Pflaster Produkte zur Infusion und Dialyse, sowie Labordiagnostika, Hilfsmittel und Medizinprodukte mit inbegriffener Software zu nennen. Die Hauptwirkung von Medizinprodukten ist in physikalischer Form vorhanden. Diese Abgrenzung zu Arzneimitteln ist für die unterschiedliche Rechtslage für den Marktzugang und der Verkehrsfähigkeit relevant [10]. Zudem sind die Medizinprodukte in drei Arten einzuteilen: (1) Die Produkte zur Labordiagnostik mit denen Tests an Proben vorgenommen werden, wie z.B. HIV-Bluttests, Schwangerschaftstests und Blutzucker-Überwachungssysteme. Diese gehören zu den In-vitro-Diagnostika. (2) Aktive implantierbare Medizinprodukte, dessen Nutzung von einer Strom- oder Energiequelle abhängig ist und (3) allgemeine Medizinprodukte [10]. Alle gemeinsam dienen dem Zweck der Erkennung, Überwachung, Kompensierung und Behandlung von Krankheiten, Behinderungen oder Verletzungen [1]. Der Einsatzort von Medizinprodukten liegt in ambulanten, stationären und alltäglichen Umgebungen [10].

2.2 Konformitätsbewertung

Für das Inverkehrbringen von Medizinprodukten sind eine Reihe an regulatorischen Erfordernisse notwendig, welche die Hersteller hinsichtlich den Medizinprodukten und Unternehmen erfüllen müssen. Um die Erfordernisse der MDR zu überprüfen, wird eine Konformitätsbewertung durchgeführt. Dieses Verfahren stellt fest, ob die Anforderungen der MDR an das Medizinprodukt eingehalten werden [1]. Mit dem Nachweis alle Anforderungen der MDR inklusive den Sicherheits- und Leistungsanforderungen zu erfüllen, gilt die Konformitätsbewertung als erfolgreich. Diese Bestätigung erlaubt die Marktzulassung der Medizinprodukte wie in Abschnitt 2.3 beschrieben ist.

Der Ablauf für die Konformitätsbewertung findet für die Medizinproduktehersteller wie

folgt statt: Im ersten Schritt legt der Hersteller die Zweckbestimmung des Medizinprodukts fest. Im Rahmen dessen wird der medizinische Zweck bestimmt. Anschließend ist anhand der oben beschriebenen Produktart die dazugehörige europäische Verordnung zu ermitteln. Medizinprodukte werden hinsichtlich der Verletzbarkeit des menschlichen Körpers in Abhängigkeit der Produktart eingestuft. Es gibt vier Risikoklassen, I, IIa, IIb und III. Die Medizinprodukte werden im Einzelfall und vom Hersteller in Bezug auf die Zweckbestimmung eigenständig einer dieser Risikoklasse zugeordnet. Je höher die Risikoeinstufung erfolgt desto detaillierter fällt das Konformitätsbewertungsverfahren aus. Zudem kann die Risikoklasse des Produktes die Einschaltung einer Benannten Stelle fordern [11]. Im nächsten Schritt wird neben dem Detaillierungsgrad auch einer der drei Arten des Konformitätsbewertungsverfahrens bestimmt.

Diese Arbeit fokussiert auf die Risikoklassen, für welche die Konformitätsbewertung auf der Grundlage eines QMS und einer Bewertung der technischen Dokumentation der MDR [1] durchzuführen ist. Die technische Dokumentation ist die Zusammenstellung aller relevanten Dokumente eines Produkts [3]. Das genannte Verfahren zur Konformitätsbewertung ist im Anhang IX der MDR geregelt. Für Medizinprodukte, die Software enthalten oder der Klasse III zugeordnet sind, wird in der Regel auf dieses Verfahren, welches ein QMS erfordert, geprüft [1]. Die anderen zwei Verfahren orientieren sich auf die Baumusterprüfung und Produktkonformitätsprüfung [1]. Bei der Herstellung von Klasse IIa- und IIb-Produkten ist das erst genannte Verfahren in einer reduzierten Form erforderlich. Für alle drei Risikoklassen gibt es zudem Sonderregelungen, welche nicht berücksichtigt werden. Hersteller von Produkten der Klasse I stellen eine EU-Konformitätserklärung gemäß Art. 19, MDR aus, wobei der Einsatz eines QMS nicht verpflichtend ist [1].

Neben der Durchführungspflicht eines Konformitätsbewertungsverfahrens, ergibt sich gemäß der MDR [1] die Erstellung eines

QMS als nächste Forderung [11]. Dies wird in Kapitel 3 näher beschrieben.

2.3 Marktzulassung

Sofern ein QMS gemäß MDR [1] im Rahmen des Marktzulassungsprozesses verwendet wird und die Produktherstellung mit der technischen Dokumentation danach ausgerichtet ist, muss der Hersteller bei der *Benannten Stelle* eine Bewertung des QMS beantragen. Die *Benannte Stelle* führt eine Überprüfung in Form eines Audits des QMS durch und prüft, ob es den gegebenen Anforderungen entspricht [11]. In dem Fall, dass das Audit bestanden ist, wird eine EU-Qualitätsmanagementbescheinigung durch die *Benannte Stelle* für den Medizinproduktehersteller ausgestellt. Damit ist die Konformität auf der Grundlage eines QMSs und einer Bewertung der technischen Dokumentation des Medizinproduktes bestätigt. Die Ausstellung eines Capillary Electrophoresis (CE)-Kennzeichens und der Eintritt für das Inverkehrbringen der Produkte in den europäischen Markt ist damit ermöglicht [10], [11]. Wenn die *Benannte Stelle* parallel die Überprüfung einer weiteren Norm vornimmt, wie z.B. DIN EN ISO 13485 [2], weisen Medizinproduktehersteller zwei Aspekte nach: (1) ein konformes QMS (2) die Bestätigung, dass das medizinische Produkt den grundlegenden Anforderungen in Bezug auf die Qualitätssicherung bzw. die Qualität genügt [11].

Mit der Betrachtung der Marktzulassung wird deutlich, dass mithilfe von Gesetzen und Verordnungen der richtige Umgang sowie die Nutzung von Medizinprodukten angestrebt wird. Damit wird ein hoher Schutz für Patienten und Anwender sichergestellt [11].

3 Qualitätsmanagement

Ein Qualitätsmanagement (QM) umfasst alle Tätigkeiten und jeweiligen Zuständigkeiten, welche dazu beitragen die Qualitätsziele einer Organisation zu erreichen [12]. Neben der Erreichung der Ziele wird im Rahmen des

QMS gemäß DIN EN ISO 9001 Verbesserungsmaßnahmen eines Produktes oder Prozesses angestrebt [5].

In der Medizinprodukte-Branche richten sich die Qualitätsziele auf die Qualitätssicherung der Medizinprodukte aus. Allgemein sind Qualitätsziele konkret und messbar festzulegen. In der Medizintechnik können Qualitätsziele in allen Unternehmensbereichen bestimmt werden. Zum einen beziehen sich die Ziele auf die Medizinprodukte und dessen Reklamations- und Fehlerquoten. Zum anderen richten sich Qualitätsziele nach den Prozessen, indem Prozesskosten oder Durchlaufzeiten analysiert werden. Mithilfe diesen Qualitätszielen wird gemäß DIN EN ISO 13485 [2] die Konformität und Zertifizierung des QMS, sowie die Konformität der Medizinprodukte aufrechterhalten [1].

3.1 QMS

Für das Herstellen und Inverkehrbringen von Medizinprodukten schreiben die Konformitätsbewertungsverfahren der MDR die Einhaltung der grundlegenden Sicherheits- und Leistungsanforderungen vor. Um die in Kapitel 4 beschriebenen Anforderungen für Medizinprodukte im gesamten Produktzyklus zu erreichen, müssen geeignete Maßnahmen organisiert und überwacht werden.

Die Aufgabe des QMS ist es einerseits die geforderten Qualitätsziele und definierten Anforderungen zu dokumentieren, verwalten und andererseits festzulegen, mithilfe welcher Maßnahmen und Prozesse diese erfüllt werden können. Es handelt sich hierbei um ein Organisations- und Managementsystem, welches gemäß DIN EN ISO 9001 [5] zum Leiten und Lenken einer Organisation hinsichtlich der Qualität eingesetzt wird [1]. Ziel dabei ist es standardisierte Prozesse und Maßnahmen auszuarbeiten, sodass Medizinprodukte nach den rechtlichen Vorgaben entwickelt und produziert werden können [12]. Sofern dies in den regelmäßig stattfindenden Audits gewährleistet werden kann, wird neben der Vorgabenkonformität langfristig die Senkung der Kosten und personellen Ressourcen im Unternehmen angestrebt [11].

Der Umfang und Detaillierungsgrad des QMS ist an dieser Stelle auch von der Risikoklasse des Medizinprodukts abhängig. Dennoch werden QMS neben der rechtlichen Pflicht auch von Unternehmen eingesetzt, um Produkte nach den Kunden- und Wettbewerbsanforderungen zu realisieren [11] [12]. Die organisatorische Gestaltung eines QMS lassen trotz den Regularien Freiräume zu. Der Hersteller entscheidet, ob die Tätigkeiten im QMS mithilfe eines Teams in einer zentralen unternehmensweiten QM-Abteilung ausgeübt oder die einzelnen Aufgaben an alle Führungskräfte oder Abteilungen delegiert werden [12]. Wenn sich der Hersteller für die erste genannte Variante, eine zentrale QM-Abteilung entscheidet, umfassen die Tätigkeitsbereiche mindestens die in Tabelle 1 dargestellten Aufgaben. Mithilfe dieses Ausschnitts an Tätigkeiten wird deutlich welcher Umfang an Aufgaben für die Umsetzung der QMS-Rolle notwendig ist. Diese ergeben sich vor allem aus den Anforderungen nach DIN EN ISO 13485 [2].

Alle aufgeführten Tätigkeiten, ID 1-6 beziehen sich auf die Sicherstellung und Gewährleistung der Medizinprodukte im gesamten Produktzyklus. Mit den Tätigkeitsfeldern ID4, ID3 und ID6 wird die stetige Verbesserung der Systemqualität, Prozessqualität und die Produktqualität angestrebt. Weitere Funktionen des QMS ergeben sich aus den Anforderungen der DIN EN ISO 13485 [2] und der MDR in Kapitel 4 [1]. Dazu gehören das in Abschnitt 2.2 vorgestellte Auditmanagement, ID 1, und die Schulung der Mitarbeiter, das in ID 6 festgehalten ist.

Neben den Verantwortlichkeitsbereichen in Tabelle 1 des QMS liegt der Fokus generell in der regelmäßigen lückenlosen Erstellung von Dokumentation und Überwachung der Prozess-, Projekt- und Änderungstätigkeiten. Deshalb wird die Tätigkeit, ID 2 Dokumentenmanagement, als Beispiel im nächsten Abschnitt näher betrachtet.

Tabelle 1 QMS- Tätigkeitsbereiche nach DIN EN ISO 13485 [2] und MDR [1]

ID	Tätigkeitsfeld	Aufgabe	Beispiele
1	Audit-Management	Vorbereitung und Durchführung von internen und externen Audits	i. Anmeldung bei Benannter Stelle ii. Vorbereitung der Dokumente iii. Durchführung des internen Audits
2	Dokumentenmanagement	Erstellung und Organisation aller Dokumente	i. Aufbewahrungsfristen ii. Dokumentenfreigabe
3	Qualitätssicherung	Erstellung der Verträge zwischen Unternehmen und Lieferanten	i. Vorgaben für Ressourcen, Methoden und Infrastruktur ii. Kommunikation iii. Überwachung der Bewertung
4	Produktion und Dienstleistung	Durchführung der Validierung der Prozesse	i. Produktions- und Dienstleistungsprozesse
5	Trainingmanagement	Mitarbeiter befähigen	i. Schulung des Personals
6	Verfahrensanweisungen (SOP)	Erstellung von Anleitungen und Prozessbeschreibungen zu rechts stehenden Beispielen	i. Produktentwicklung ii. Messung, und Verbesserungen der Produkte iii. Validierung von Prozessen und Computer Software v. Risikomanagement

3.2 Dokumentenmanagement

Aus der Tätigkeitsbeschreibung ID 1 bis 6 in Tabelle 1 des QMS geht hervor, dass in allen Aufgaben Dokumente erzeugt, überprüft oder Ergebnisse und Informationen festgehalten werden. Aufgrund dessen ist ein Dokumentenmanagement mit der Lenkung von Dokumenten eines der Hauptsäulen eines

funktionierenden QMS für Medizinproduktehersteller. Unter der Dokumenten-Lenkung ist das kontinuierliche Aktualisieren und die gezielte Verteilung der Dokumenteninhalte in einem Unternehmen zu verstehen.

Neben allgemeinen Unternehmens- und Qualitätszielen fallen Vorgabe- und Nachweisdokumente unter die Pflicht der Hersteller. Dabei verpflichten die DIN EN ISO 13485 [2] und DIN EN ISO 9001 [5] die Medizinproduktehersteller die Erstellung, Freigabe, Verteilung und Aktualisierung von Vorgabedokumenten- und Nachweisdokumente sowie Qualitätsnachweisen nachzugehen. Beispiele sind der Abbildung 1 zu entnehmen. Vorgabedokumente sind verbindlich. Als Basis dient ein Qualitätshandbuch, siehe (1) in Abbildung 1, das den Umfang und die Ziele des QMS mit den jeweiligen Prozessen in der Organisation beschreibt. Prozesse, welche im Rahmen des QMS definiert werden, geben vor wie Vorgänge im Unternehmen durchgeführt werden müssen und wer dafür verantwortlich ist, siehe (2). Diese Vorgehensweisen beziehen sich beispielweise darauf wie das Projektmanagement, das Produkt-Risikomanagement oder die Überwachung der Medizinprodukte nach dem Inverkehrbringen erfolgen muss.

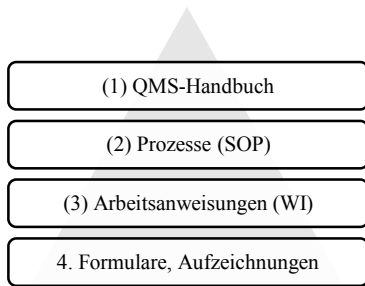


Abbildung 1 Vorgabe- und Nachweisdokumente

Mit der Definition von Prozessvorgehensweisen, im engl. als *Standard Operating Procedure* (SOP), wird die Erfüllung von Normen für die unternehmensbezogenen Vorgänge und Prozesse erleichtert, siehe (2).

Weitere zu erstellende und pflegende Dokumente sind Arbeitsanweisungen, im engl. als *working instruction* (WI), siehe (3) bezeichnet [8]. Diese beschreiben bestimmte Aufgaben z.B. wie die Verpackung von Produkten erfolgen muss. Die Aufzeichnungen gelten als Nachweisdokumente und unterscheiden sich von den Vorgabedokumenten, indem diese beweisen welche Tätigkeit umgesetzt wurde. Ein Beispiel ist der Nachweis einer Schulung [9].

Neben der Erstellungspflicht dieser Dokumente wird explizit die Dokumentenfreigabe und Prüfung in Form der Dokumenten-Lenkung geregelt. Diese Lenkung von Dokumenten regelt, dass die Dokumente von unterschiedlichen Personen erstellt, geprüft und freigegeben werden müssen [9]. Die Kombination aus Dokumentenerstellung, die Dokumenten-Lenkung und Schulungsmanagement als Teil der Funktion des QMS wird anhand eines Beispiels in Abbildung 2 veranschaulicht.

In Abbildung 2 werden an zwei verschiedenen Vorgabedokumenten, SOP und WI, ein theoretisch möglicher Ablauf im Rahmen des Dokumentenmanagementprozesses gezeigt. Der Ablauf orientiert sich am verallgemeinerten Verlauf einer Erstellung von Prozessen und Arbeitsanweisungen in regulierten Branchen wie der Medizintechnik. Beispiele sind der ID 6 in Tabelle 1 zu entnehmen. Nach der Bestimmung der Art und Sprache des Dokuments ist die entsprechende Vorlage zu verwenden. Dies wird in der Regel von einem Prozessverantwortlichen durchgeführt. Es ist möglich in Zusammenarbeit das Dokument zu erstellen. Dabei ist es wichtig die Inhalte für die Anwender präzise, verständlich und mittels Abbildungen zu formulieren. Unabhängig davon ist mindestens das Vier-Augen-Prinzip anzuwenden, sodass jeweils ein Prüfer und ein Genehmiger das Dokument akzeptieren müssen. Nach der Entscheidung über eine Schulung oder Informationsausgabe tritt das Dokument in Kraft und ist damit gültig. Sofern eine Änderung und Anpassung ansteht wiederholt sich der Vorgang ab dem Punkt *Zusammenarbeit*.

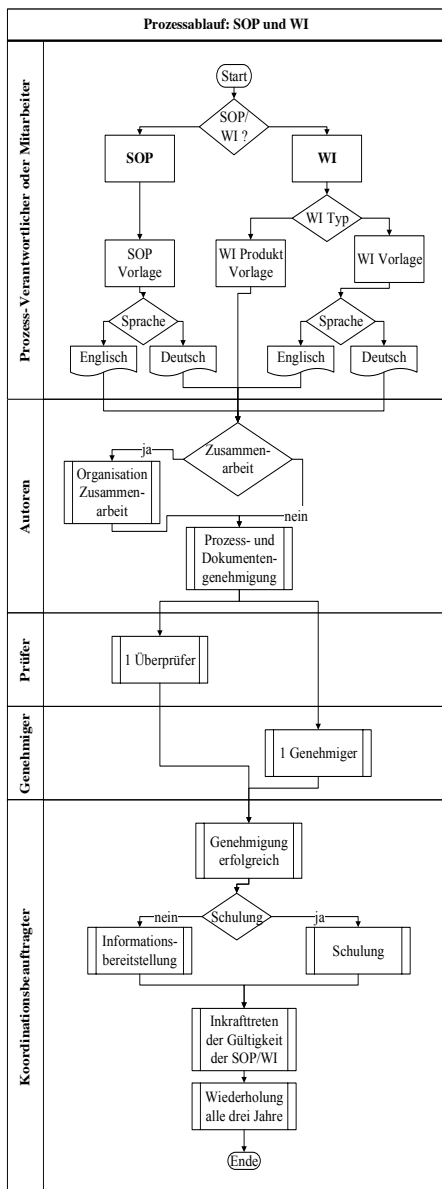


Abbildung 2 Vorgabedokumente QMS

3.3 Elektronisches QMS

Mit der Pflicht ein QMS, alle Prozeduren, Verantwortlichkeiten und Dokumente aufzu-

bauen und organisieren, unterliegen Medizinproduktehersteller einem zeitlichen und personellen Aufwand. Langfristig gesehen will der Gesetzgeber damit Einsparungen erbringen, in dem z.B. die Entwicklungszeiten von Produkten verkürzt und Prozesse optimiert werden. Andererseits sollen Rollen und Berechtigungen zwischen den Mitarbeitern eindeutig geregelt sein [8]. Angelehnt an den bestehenden Umfang und Aufwand der QMS-Tätigkeiten setzten in Zeiten der Digitalisierung immer mehr Hersteller Softwarelösungen ein. Diese werden nach aktuellen Stand der Technik in der Regel als funktionsübergreifende QMS-Lösung eingesetzt und damit als eQMS zu bezeichnen. Dabei sollen die in Tabelle 1 zusehenden Tätigkeitsfelder unterstützt werden [3] [5].

Die Digitalisierung des Dokumentenmanagementprozesses, wie am Beispiel in Abbildung 2 zu sehen, kann dazu beitragen, die gesamte Prozessdauer von bis zu mehreren Wochen zu reduzieren. Dies erfolgt mit der Nutzung eines eQMS, indem Prozessabläufe automatisiert werden und elektronische Unterschriften für die Dokumentengenehmigung eingesetzt werden können. Außerdem ist es mit einem eQMS möglich tägliche Aufgaben sowie den Zertifizierungsprozess zu vereinfachen und effizienter zu gestalten.

Wenn Softwarelösungen im Kontext des QMS eingesetzt werden, sind diese nicht als Medizinprodukt definiert. Die Verwendung als Medizinprodukt wird ausgeschlossen, da keine elektronische oder manuelle Vernetzung mit den Produkten besteht. Zudem werden keine Daten durch ein Medizinprodukt erzeugt und nicht zum therapeutischen oder diagnostischem Kontext eingesetzt [13]. Dennoch muss im nächsten Kapitel analysiert werden, was sich für den Medizinproduktehersteller bei der Einführung und Nutzung einer solchen Software hinsichtlich den regulatorischen Vorgaben ändert. Im Rahmen des QMS müssen weiterhin alle Tätigkeiten wie in Tabelle 1 zu sehen, normgerecht ausgeführt werden.

4 Normen und Gesetze für ein QMS

In den vorherigen Kapiteln wurde festgehalten, dass Gesetze und Normen Anforderungen an ein QMS und Medizinprodukte stellen. In diese Kapitel wird verdeutlicht welche Regularien und welche Vorgaben das sind.

4.1 Regularien an ein QMS

Für die Betrachtung der regulatorischen Anforderungen an ein QMS erfordert es den Einbezug des Kontextes, da diese Vorgaben Auswirkungen auf ein QMS nehmen. Deshalb sind Vorgaben für Medizinprodukte, ein allgemeines Qualitätsmanagement sowie für ein eQMS zu betrachten [13]. Die Untersuchung der jeweiligen Gesetzeslage in Abbildung 3 ist deshalb erforderlich. Für Medizinprodukte, welche als aktiv implantierbare und In-vitro Diagnostik definiert werden, gibt es jeweils eine eigene Verordnung. Diese werden in dieser Arbeit nicht untersucht.

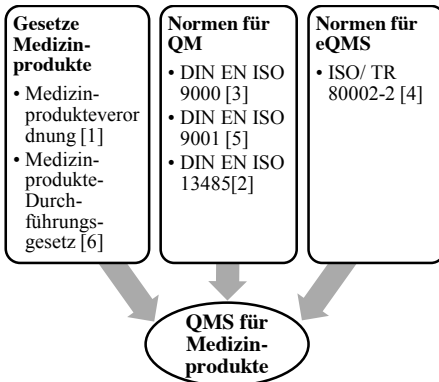


Abbildung 3 QMS bezogene Regularien

Mit der Veränderung der europäischen Rechtslage ist die Medizinprodukteverordnung [1] und das MPDG [6] als übergreifendes-Recht ab dem 26.05.2020 zu betrachten. Als europäisches Recht sind diese für alle Medizinproduktehersteller der Europäischen Union (EU) verpflichtend. Die bisher bestehenden Richtlinien sind damit nicht mehr gültig. Die MPDG ergänzt die MDR [1] [12].

Die Normen in Abbildung 3 sind als EU-Empfehlung an den Stand der Technik angelehnt. Diese decken die Themengebiete des allgemeinen QM nach DIN EN ISO 9000 [3] und 9001 [4] sowie die Erfordernisse für ein umfassendes QMS für die Herstellung von Medizinprodukten nach DIN EN ISO 13485 [2] ab [7]. Die ISO TR 80002-2 [4] beschreibt als *Technischen Report* (Technical Report, TR) die Anforderungen an die Validierung von Software, die im Rahmen von QMS-relevanten Prozessen eingesetzt wird. Damit können die Anforderungen der DIN EN ISO 13485 nach Prozess-Software-Validierung analysiert werden [2] [13].

Mit der Betrachtung der aufgeführten Regularien werden zunächst die Anforderungen des QMS analysiert. Im Anschluss wird untersucht, ob diese mit der Einführung eines eQMS bestehen bleiben oder ob sich Änderungen für den Medizinproduktehersteller und dessen QMS ergeben.

4.2 Regulatorische Anforderungen an ein QMS

In diesem Abschnitt sind die herausgearbeiteten Anforderungen an ein QMS anhand den relevanten Gesetzen und Normen aus Abbildung 3 in Tabelle 2 dargestellt. Die sind den Tätigkeitsfeldern in der ersten Spalte zugeordnet. Wobei die Anforderungen in der mittleren und die Aufgaben zur Umsetzung in der rechten Spalte dokumentiert sind.

Insgesamt ist festzuhalten ist, dass die Regularien vor allem Anforderungen an das Dokumentenmanagement und -Lenkung, Prozessmanagement hinsichtlich der Produktion, das Bereitstellen der notwendigen Ressourcen und Methoden sowie die stetige Kontrolle und Überwachung der Verfahren- und Nachweisdokumente stellen. Aus den Anforderungen ergeben sich die erforderlichen Aufgaben, welche ein QMS umsetzen und erfüllen muss. Ob sich diese Anforderungen bei der Einführung und Nutzung einer QMS unterstützenden Software ändern und ob ein QMS angepasst werden muss ergibt die nächste Analyse der Normen im Abschnitt 4.3.

Tabelle 2 Anforderungen an ein QMS nach ISO 13485 [2], MPDG [6], MDR [1]

Tätigkeitsfeld	Anforderungen	Umsetzung der Anforderung
Produkt-Lebenszyklus	Struktur, Verantwortlichkeiten, Prozesse und Managementressourcen	Planung und Dokumentation der Ressourcen der Produktrealisierung
	Überwachung des gesamten Produktlebenszyklus mit und ohne Software	Überwachung von Daten über die Qualität, Leistung und Sicherheit der Produkte sowie Abklärung hinsichtlich der Software
	Produktrealisierung, Produktentwicklung und zur Auslieferung, Installation und Instandhaltung	Durchführung, Überwachung des Entwicklungsprozesses mit Validierung, Verifizierung von Medizinprodukten mit Software.
	Inverkehrbringen, Prozessdefinition, Unique Device Identifikation (UDI)-Kennzeichnung von Medizinprodukten	Rückverfolgbarkeit, Meldepflicht an UDI-Datenbank, Datenbankeintrag
	Kontrolle: Zulieferer und Lieferanten	Ressourcenmanagement, Lieferantenkontrolle
Sicherheit	Einhaltung Sicherheits- und Leistungsanforderungen	Entsorgung, IT-Sicherheit, Kennzeichnung
Allgemeine Anforderungen	Rollen- und Berechtigungskonzepte	Festlegung der Zuständigkeiten des Personals hinsichtlich Tätigkeit und Prozesse
	Prozesse zur Umsetzung der rechtlichen Anforderungen	Definition von unternehmensinternen Prozessen in Form von SOPs
	Verfahren, Tätigkeiten, Regelungen erstellen und aufrechterhalten	Alle SOPs, WIs, Aufzeichnungen anhand Regularien erstellen
	Gebrauchstauglichkeit und Zweckbestimmung von Patienten, Anwendern	Format der Gebrauchsanweisung, Lesbarkeit, Verständlichkeit, Verfügbarkeit
	Risikomanagement	Prävention und Beseitigung von Risiken im Risiko-Nutzen-Verhältnis
Verantwortung der Leitung, QM-Beauftragter und Hersteller	Qualitätsmanagement ist die Aufgabe der Unternehmensleitung oder ein Beauftragter der obersten Leitung Qualitätsziele und Qualitätspolitik	Startimpuls für Einführung und Aufrechterhaltung eines QMS nach Stand der Technik und Überprüfung
	QMS ab Risikoklasse III verpflichtend	Überprüfung der Risikoklasse und Art des Produktes, Überprüfung der Rechtslage
Dokumentenmanagement	Erstellung, Freigabe, Lenkung und Aktualisierung von Dokumenten	Dokumentation aller Vorgabe- und Nachweisdokumenten, Aufbewahrungspflicht
	Sicherstellung eines geeigneten QMS mittels Dokumentation	Umsetzung der Anforderungen, Aufbau, Umsetzung, Aufrechterhalten und Verbesserung eines QMS mit Festhaltung der Änderungen
	Technische Dokumentation für alle Medizinprodukte	Erstellung der technischen Dokumentation für den gesamten Produktlebenszyklus
Schulungen	Mitarbeiterqualifizierung, Rollen- und Berechtigungskonzept	Mitarbeiter für ihre Rollen und Verantwortlichkeiten qualifizieren.
Infrastruktur	Gestaltung der Infrastruktur und Arbeitsumgebung des Unternehmens	Mitgestaltung der Räumlichkeiten, Maschinen, Werkzeuge oder Umgebungsbedingungen wie Luftfeuchte, Temperatur
Prozessmanagement	Definition und Steuerung aller QMS Prozesse nach Abschätzung des Risikos und Ressourcen als Prozess-Lenkung	Durchführung aller Prozesse mit Dokumenten anhand des Planen-Durchführen-Prüfen-Handeln – Zyklus (PDCA), Methodenimplementierung
Audit Management	Informationspflicht gegenüber <i>Benannter Stelle</i>	Nachweis einer vollständigen Dokumentation Austausch mit <i>Benannter Stelle</i>
Messung, und Verbesserung	Sicherstellung des QMS, Auslegungs- und Produktqualität, Verbesserungsstrategien	Überwachungs-, Korrekturmethode eines effizienten QMS

4.3 Regulatorische Anforderungen an ein eQMS

Die Anforderungen an ein QMS sind mithilfe der zuvor durchgeführten Analyse in Abschnitt 4.2 bekannt. Im nächsten Schritt wird überprüft welche Änderungen und Anpassungen bei der Einführung einer Software für das QM anstehen. Ziel ist es weiterhin ein normgerechtes QMS für die Auditierung und Zertifizierung durch *Benannte Stellen* zur Verfügung stellen zu können. Grundsätzlich müssen mindestens alle in Tabelle 2 gelisteten Anforderungen eingehalten werden. Dies betrifft auch die Anforderungen hinsichtlich des Produktzyklus.

Die MDR beschreibt konkret welche Maßnahmen bei Veränderungen gemäß der Informationspflicht notwendig sind. Demnach unterliegen alle Medizinproduktehersteller der Forderung, Änderungen bereits in der Planungsphase der *Benannten Stelle* mitzuteilen. Dies beinhaltet auch die Modifikation am bereits genehmigten QMS. Als Beispiel kann die Einführung eines eQMS und dessen Änderungen hinsichtlich der Lenkung und Genehmigen von Dokumenten genannt werden. Die *Benannte Stelle* bewertet die Vorschläge und entscheidet ob zusätzliche Audits als Zertifizierung erforderlich sind. Unabhängig der Entscheidung wird ein QMS aufgrund der Veränderung nach seiner Wirksamkeit überprüft. Mit dem Nachweis eines effizienten QMS wird die Information als Nachtrag zur EU QM-Bescheinigung ergänzt [1].

Der erfolgreichen Überprüfung ist die Software-Validierung vorangestellt. Die DIN EN ISO 13485 verpflichtet Medizinproduktehersteller Software, die im Rahmen des QMSs eingesetzt wird, zu validieren [2]. Die Validierung muss vor dem ersten Einsatz durchgeführt werden und das gesamte Verfahren ist zu dokumentieren. Unter der Validierung ist die Prüfung der Eignung einer Software in Bezug auf den Einsatzzweck zu verstehen. Ein Beispiel für ein eQMS ist eine Software zum Verwalten von Kundenrückmeldungen, eine Software als Dokumentenmanagement

oder die Kombination von Aufgaben aus verschiedenen Tätigkeitsbereichen des QMS. Sofern Änderungen wie beispielsweise das Hinzufügen von Prozessen oder weitere Tätigkeitsbereiche wie das Schulungsmanagement konfiguriert werden, ist eine weitere Validierung verpflichtend. Das Risiko muss bei jeder Validierung berücksichtigt werden. Das betrifft alle Softwarelösungen, die im Rahmen eines Prozesses, welche Teile des QMS steuert, verwendet werden [2] [8].

Normbasierte Anforderungen wie die Einführung von Software im QMS werden explizit nicht von der MDR [1] und DIN EN ISO 13485 [2] geäußert. Die Anforderungen beziehen sich auf den Stand vor der Einführung. Zusammengefasst ergeben sich folgende Pflichten für ein QMS und dessen Unternehmen:

- (1) QMS bezogene Anforderungen aus der MDR [1] und der DIN ISO 13485 [2] bleiben weiterbestehen und sind sicherzustellen
- (2) Während der Planung eines eQMS ist die *Benannte Stelle* zu informieren
- (3) Definition des zukünftig zu unterstützenden Aufgabenbereich des QMS
- (4) Planung, Durchführung der Software Validierung
- (5) Bei Änderung des eQMS Informationspflicht gegenüber *Benannter Stelle*
- (6) Einführung, Entwicklung der Software auf Regularien an ein QMS abstimmen.

Für das Dokumentationsmanagement bedeutet das, dass mit einem eQMS weiter die Anforderungen der Dokumenten-Lenkung sowie der Archivierungsfristen eingehalten werden müssen. Idealerweise werden die Dokumentenprüfungen in Form von elektronische Prozessschritten integriert. Mithilfe einer Software können beispielsweise auch Vor-

gabe- und Nachweisdokumente zum jeweiligen Produkt erstellt oder zumindest vorbereitet werden.

Im nächsten Kapitel wird skizziert wie Medizinproduktehersteller die Anforderungen der DIN ISO 13485 [2] im Hinblick auf die Prozess-Software-Validierung erfüllen können.

5 Software-Validierung eines eQMS

In Kapitel 4 konnten die Pflichten bei der Planung und Einführung einer Software für ein QMS festgehalten werden. Im weiteren Verlauf wird verdeutlicht, ob und welche Vorgaben an die Prozess-Software-Validierung vorliegen [4].

Der *Technische Bericht* ISO TR 80002-2 [4] beinhaltet die Anforderungen dieser Software-Validierung, welche die DIN EN ISO 13485 [2] fordert. In den Anwendungsbereich dieser Norm fällt jede Software, die im Rahmen des QMS, in der Produktion oder bei der Erbringung von Dienstleistungen genutzt wird. Zum Beispiel voll- und teil automatisierte Prozesse zum Umgang mit Beschwerden, Entwicklung oder der Lenkung von Dokumenten [2]. Die ISO TR 8000-2 [4] soll als *Best Practice* und als Empfehlung betrachtet werden. Dennoch wird sie in dieser Arbeit untersucht, um Informationen für die Einführung einer Software ins QMS zu gewinnen. Die Norm betrachtet den gesamten Software-Lebenszyklus und fordert die stetige Betrachtung der Risiken, welche sich bei einer fehlerhaften Software auf das QMS auswirken können [4]. Zu Beginn ist festzulegen welche Anforderungen das Unternehmen an die Prozesse, die mit der Software unterstützt und automatisiert werden sollen, festlegt. Die Risiken für Prozessfehler sind zu beachten. Im Anschluss kann der Entwurf eines Validierungsplans erstellt werden. Dieser berücksichtigt unter anderem die zuvor definierten Prozessanforderungen. Mit der Zweckbestimmung wird festgelegt wer zukünftig die Software nutzt und welche Vorgaben diese stellen. Darüber hinaus sind Softwareanforderungen zu definieren. Anschließend wird

der Validierungsplan überarbeitet. In Abstimmung mit der *Benannten Stelle* wird die Software entwickelt. Nach der Entwicklung wird empfohlen eine Testumgebung vorzubereiten, in welcher Testfälle spezifiziert werden. Die zuvor definierten Software- und Nutzeranforderungen werden damit schrittweise überprüft und dokumentiert [2].

6 Fazit und Ausblick

Ziel dieser Arbeit ist es mithilfe den regulatorischen Anforderungen wie der MDR [1] und DIN EN ISO 13485 [2] herauszuarbeiten, welche Vorgaben bei der Einführung der eQMS zu beachten sind und wie diese erfüllt werden können. Damit wird festgestellt, dass sich die Anforderungen an ein QMS, wie in den Gesetzten bereits vorgenommen, in Tätigkeitsbereiche einteilen lassen. Die Analyse der Anforderungen und Beschreibung von Aufgaben einzelner Tätigkeitsfelder des QMS, bietet die Grundlage dafür welche Vorgaben im Rahmen einer Softwareeinführung notwendig sind. Für die Einbindung eines eQMS sind wenige, jedoch klare Anforderungen seitens der DIN EN ISO 13485 und der MDR gefordert. Die Hauptanforderung liegt eindeutig in der Software-Validierung, welche vor dem Einführen der Software verpflichtend umzusetzen ist. Ob ein QMS mit der Einführung einer Software weiterhin als gültig und effizient einzustufen ist, prüft die *Benannte Stelle*. Da sich Medizinproduktehersteller Fehler und Probleme eines QMS nicht erlauben dürfen, ist es ratsam sich den *Technischen Bericht* der ISO TR 8000-2 [4] anzusehen. Dort ist der Ablauf beschrieben, welche für die Planung und Umsetzung einer Software relevant ist. Dieser Bericht kann als Empfehlung für Projekte in der Praxis verwendet werden, welche ein eQMS einführen wollen. Die wichtigsten Schritte wurden dabei in dieser Arbeit zusammengefasst.

Ausblickend ist festzuhalten, dass für die praktische Umsetzung und Einführung einer Software für ein QMS die theoretische Anleitung hilfreich, jedoch nicht vollständig ist. Zudem muss beachtet werden, dass das allgemeine QMS anhand den Risikoklassen und

Produktarten ausgerichtet ist, welche auch bei der Nutzung von eQMS relevant werden. Demnach wird eine Untersuchung der spezifischen Verordnungen und Anforderungen der jeweiligen Medizinproduktart und -klasse in zukünftigen Arbeiten empfohlen. Insgesamt ist festzuhalten, dass mit dieser Arbeit das grundlegende Verständnis sowie die systematische Analyse der Anforderungen an ein QMS und der Einführung an ein eQMS erarbeitet werden konnte.

7 Literaturverzeichnis

- [1] Medical Device Regulation MDR – Medizinprodukteverordnung (2017/745) Verordnung (EU) 2017/745 des Europäischen Parlaments und des Rates vom 5. April 2017 über Medizinprodukte, zur Änderung der Richtlinie 2001/83/EG, der Verordnung (EG) Nr. 178/2002 und der Verordnung (EG) Nr. 1223/2009 und zur Aufhebung der Richtlinien 90/385/EWG und 93/42/EWG des Rates.
- [2] DIN EN ISO 13485:2016-08 Medizinprodukte - Qualitätsmanagementsysteme - Anforderungen für regulatorische Zwecke (ISO 13485:2016); Deutsche Fassung EN ISO 13485:2016-08.
- [3] DIN EN ISO 9000:2015-11 Qualitätsmanagementsysteme - Grundlagen und Begriffe (ISO 9000:2015); Deutsche und Englische Fassung EN ISO 9000:2015-11.
- [4] ISO/TR 80002-2:2017-06 Medizinische Gerätesoftware - Teil 2: Validierung von Software zur Verwendung in der Qualitätssicherung für medizinische Geräte.
- [5] DIN EN ISO 9001:2015-11 Qualitätsmanagementsysteme - Anforderungen (ISO 9001:2015); Deutsche und Englische Fassung EN ISO 9001:2015-11.
- [6] Medizin-Durchführungsgesetz 19/15620 Gesetzentwurf der Bundesregierung. Entwurf eines Gesetzes zur Anpassung des Medizinprodukterechts an die Verordnung (EU) 2017/745 und die Verordnung (EU) 2017/746.
- [7] Gärtner, A (2017) Kommunizierende medizinische Systeme und Netzwerke In: Kramme R. (eds) Medizintechnik. Springer Reference Technik. Springer, Berlin, Heidelberg.
- [8] Medizinprodukte - Qualitätsmanagementsysteme - Anforderungen für regulatorische Zwecke (ISO 13485:2016); Deutsche Fassung EN ISO 13485:2016Bijan, E., (2018) Safety Risk Management for Medical Devices. London, Academic Press an imprint of Elsevier.
- [9] Harer, J., & Baumgartner, C. (2018). Anforderungen an Medizinprodukte: Praxisleitfaden für Hersteller und Zulieferer. Carl Hanser Verlag GmbH & Co. KG.
- [10] Lauer, W., Stöblein, E., Brinker, A., & Broich, K. (2014). Medizinprodukte. Bundesgesundheitsblatt-Gesundheitsforschung-Gesundheitsschutz, 57(12), 1355-1361.
- [11] Malkmus A. (2017) Qualitätsmanagement in der Medizintechnik: Ziele, Elemente und Strukturen. In: Kramme R. (eds) Medizintechnik. Springer Reference Technik. Springer, Berlin, Heidelberg.
- [12] Prinz, T. (2017). Entwicklung und Herstellung medizinischer Software. VDE-Schriftenreihe Normen verständlich, (171).
- [13] Manz, S. (2018). Medical Device Quality Management Systems: Strategy and Techniques for Improving Efficiency and Effectiveness. London, Academic Press an imprint of Elsevier.

Analyse des Requirements Engineering eines KIS-Herstellers

Katharina Lohr

Hochschule Reutlingen

Katharina_Barbara.Lohr@Student.Reutlingen-University.de

Abstract

Requirements Engineering (RE) umfasst sämtliche systematische Schritte zur Entwicklung eines Systems, um die Bedürfnisse der Nutzer und Vorgaben, die an dieses gestellt werden zu erfüllen. Das RE eines ausgewählten Herstellers für klinische Informationssysteme (KIS) wurde untersucht und es stellt sich als intransparent als auch teilweise unzureichend dar. Das Ausmaß des Einsatzes von systematischen Vorgehensweisen und Methoden zum RE wurden beim ausgewählten KIS-Hersteller analysiert. Die Analyse zeigt, dass RE weit verbreitet ist, aber differenziert betrieben wird. Das Ziel dieser Arbeit ist es, den Stand der Technik des RE für die KIS-Entwicklung zu ermitteln. Es werden wichtige Faktoren des RE für die Entwicklung von KIS beschrieben. Die Ergebnisse dieser Arbeit werden als erster Schritt für die Optimierung des RE des ausgewählten KIS-Herstellers dienen.

Schlüsselwörter

Requirements Engineering, Klinische Informationssysteme, Medizinische Informationssysteme, Goal Question Metric

CR-Kategorien

•Information systems ~ Information systems applications

Betreuer Hochschule: Prof. Dr. Christian Kücherer
Hochschule Reutlingen
christian.kuecherer@reutlingen-university.de

Wissenschaftliche Vertiefungskonferenz
20. Mai 2020, Hochschule Reutlingen
Copyright 2020 Katharina Lohr

1 Einleitung

1.1 Relevanz und Kontext

Die Durchführung des Requirements Engineerings (RE) ist für die erfolgreiche Entwicklung von klinischen Informationssystemen (KIS) unverzichtbar [1]. RE ist das systematische Vorgehen zur Entwicklung, Umsetzung und Verwaltung von qualitativen, wertschaffenden und risikoorientierten Anforderungen im gesamten Lebenszyklus eines Produktes [2]. KISs sind Software-Lösungen für klinische Einrichtungen, die den administrativen und medizinischen Datenaustausch, über verschiedene Abteilungen, zur Dokumentations-, Prozess-, Entscheidungs- und Kommunikationsunterstützung ermöglichen [3]. In dieser Arbeit werden KISs als Verbund einzelner Systeme betrachtet, die in ihrer Gesamtheit, die Funktionalitäten eines KIS ausmachen. Als medizinische Informationssysteme (MIS) gelten in dieser Arbeit hingegen Informationssysteme, die Daten im medizinischen Kontext zur Versorgung von Patienten verarbeiten [3] und somit der Zweckerfüllung eines KIS nahekommen. In der Informatik ist das RE von zentraler Bedeutung, um qualitativ hochwertige Software zu entwickeln, welche die Nutzerbedürfnisse erfüllt [4]. Die Qualität von Softwaresystemen ist vom Erfüllungsgrad der an das System gestellten Anforderungen abhängig [5]. Anforderungen beschreiben die Bedürfnisse der Nutzer, die ein System in einer bestimmten Art und Weise unterstützen soll [6]. Nach Va-

lentini müssen zur Durchführung des RE verschiedene Herausforderungen bewältigt werden. Zu diesen zählt das Vorhandensein eines gemeinsamen Verständnisses zu Anforderungen und sich ändernden Anforderungen, um unzureichende Anforderungen zu vermeiden [6]. Nach Patton wird beim gemeinsamen Verständnis von verschiedenen Personen, zu einer bestimmten Vorstellung, das Gleiche verstanden. Dieses gemeinsame Verständnis entsteht durch Kommunikation zwischen den Beteiligten. Dafür ist die Verwendung der gleichen Sprache, das Erörtern von Ausdrücken oder das Modellieren von Informationen als bildliche Darstellung zu einer bestimmten Vorstellung, notwendig [7]. Die Vielseitigkeit von Definitionen im RE, als auch die vieldeutige Verwendbarkeit der natürlichen Sprache, gelten als häufige Ursache dieser Missverständnisse [8]. Dementsprechend kann eine Anforderung von verschiedenen Personen unterschiedlich verstanden werden. Es kann die Gefahr bestehen, dass relevante Anforderungen unzureichend aufgenommen oder nicht berücksichtigt werden [7].

1.2 Problematik und Fragen

Die Entstehung dieses Artikels basiert auf der Zusammenarbeit mit einem KIS-Hersteller, der nachfolgend als Referenz-KIS-Hersteller bezeichnet wird. Basierend auf der Experten-Erfahrung des Referenz-KIS-Herstellers wird die KIS-Entwicklung aufgrund teilweise unpräzise formulierter, unverständlicher oder sich ändernder Anforderungen erschwert. Die Folge ist eine allgemeine Intransparenz, die sowohl Kunden als auch Mitarbeitern bemängeln. Durch diese Intransparenz werden die Priorisierung und Umsetzung einzelner Anforderungen teils unmöglich. Nach Angaben des Referenz-KIS-Herstellers ist die Reife des RE bei diesem unbekannt. Die Reife des RE beschreibt in dieser Arbeit das Ausmaß des Einsatzes von systematischen Vorgehensweisen und Methoden zum RE. Die in dieser Arbeit untersuchte Problemstellung ist: Für eine erfolgreiche KIS-Entwicklung ist ein RE, dass

einheitlich verstanden, kommuniziert und qualitativ praktiziert wird, notwendig. Aus dieser Problemstellung werden die nachfolgenden Forschungsfragen abgeleitet:

F1: Welche Vorgehensweisen und Aktivitäten umfasst derzeit das RE für KIS?

F2: Können Kriterien für die Qualität des RE in der KIS-Entwicklung allgemeingültig abgeleitet werden?

F3: Welche Reife hat das RE bei einem ausgewählten Referenz-KIS-Hersteller?

1.3 Zielsetzung

Basierend auf diesen Forschungsfragen hat diese Arbeit zum Ziel, den Stand der Technik zum RE in der KIS-Entwicklung zu ermitteln. Darüber hinaus soll das Potenzial von RE für die Entwicklung von KIS aufgezeigt werden. Auf Grundlage der neu gewonnenen Erkenntnisse, soll eine Handlungsempfehlung zur Ermittlung der Reife des RE des ausgewählten Referenz-KIS-Herstellers gegeben werden.

1.4 Aufbau

Diese Arbeit ist wie folgt aufgebaut: In Kapitel 2 werden die Grundlagen der Literatur zum Verständnis und der Durchführung von RE aufgezeigt. Kapitel 3 beinhaltet eine Untersuchung des RE für KISs. In Kapitel 4 werden die Ergebnisse dieser Untersuchung präsentiert. Durch einen Vergleich der Erkenntnisse der vorangegangenen Kapitel werden in Kapitel 5 die Ergebnisse dieser Arbeit diskutiert. In Kapitel 6 ist das Fazit mit einer Handlungsempfehlung für den ausgewählten Referenz-KIS-Hersteller formuliert. Abschließend wird in einem Ausblick die künftige Verwendung der gewonnen Ergebnisse dieser Arbeit aufgezeigt.

2 Grundlagen

In diesem Kapitel werden für diese Arbeit relevante Begriffe, Vorgehensweisen, Aktivitäten und Empfehlungen der derzeitigen Literatur zum RE kurz vorgestellt.

2.1 *Qualität und Anforderung*

Nach Ebert ist Qualität bezüglich der Beschaffenheit einer Software, die vollständige Erfüllung aller Erwartungen an diese [2]. Einen anderen Ansatz beschreibt Ludewig [9], nach dem Qualität eine subjektive Eigenschaft einer Software ist, die nicht gemessen werden kann. Es können lediglich die Defizite identifiziert werden, welche durch das Fehlen von Qualität entstehen. Forderungen, die sich auf eine spezifische Qualität eines Software-Systems beziehen, werden als Anforderungen bezeichnet [9]. Sowohl in der ISO 24765 [10] als auch im Glossar des International Requirements Engineering Boards (IREB¹), wird eine Anforderung als eine Fähigkeit oder Eigenschaft eines Systems beschrieben. Dagegen unterscheidet sich die Definition der Anforderung in den zwei Werken bezüglich ihrer Zweckmäßigkeit. Nach der ISO 24765 [10] dient eine Anforderung dazu, „um ein bestimmtes Problem zu lösen oder ein bestimmtes Ziel zu erreichen“ [10] und zur Erfüllung einer Spezifikation eines Vertrags, einer Norm oder eines weiteren formellen Dokuments. Gemäß des IREB Glossars beschreiben Anforderungen, den Bedarf von Stakeholdern. Zusätzlich dienen Anforderungen der Dokumentation des Bedarfs als auch der Fähigkeit oder Eigenschaft eines Systems. Stakeholder sind „Einzelperson oder Organisationen, die ein Anrecht, einen Anteil, einen Anspruch oder ein Interesse auf ein bzw. an einem System oder an dessen Merkmalen haben, die ihren Erfordernissen und Erwartungen entsprechen“ [11].

2.2 *Ansätze des RE*

Wieringa [12] unterteilt die Vorgehensweisen zum RE in zwei Denkrichtungen: (i) dem Problemorientierten RE (PORE) und (ii) dem Lösungsorientierten RE (SORE). PORE fo-

kussiert auf eine Theorie bezüglich eines praktischen Problems in der realen Welt. Dabei ist das Verstehen der Problematik wesentlich. Es zielt nicht darauf ab das Problem durch die Veränderung des Ursprungs zu lösen. Vorgehensweisen, die auf dem Verständnis von PORE beruhen sind nach Wieringa die Techniken des Zielorientierten RE (GORE) [12].

GORE fokussiert sich auf die Bestimmung und Bedeutung von Zielen zur Durchführung der Aktivitäten des RE. Durch die Verwendung von GORE können Anforderungen durch die Erstellung von Zielmodellen, systematisch abgeleitet, verfeinert, begründet und nachvollzogen werden. Zielmodelle können durch verschiedene zielorientierte Sprachen als gerichtete Graphen mit UND/ODER Beziehungen modelliert werden [14].

SORE bezeichnet Wieringa als eine Tätigkeit des Designs, um die Differenz zwischen einem gewünschten Zustand und einem vorliegenden Zustand, durch spezifische Handlungen zu vermindern. Dafür wird zuerst eine Implementierung einer Lösung entworfen und darauf in Form einer Spezifikation der Anforderungen beschrieben [12]. Vorgehensweisen des SORE sind bspw. das Erstellen einer Spezifikation für Anforderungen gemäß der ISO 29148 [15] und der Einsatz des Volere Templates [13] von Robertson. Das Volere Template berücksichtigt funktionale und nicht funktionale Anforderungen, sowie die Erstellung von Anwendungsfallspezifikationen und Anwendungsfalldiagrammen, um Anforderungen und Einschränkungen dieser zu identifizieren [16].

In der Literatur existieren Ansätze zum RE, die PORE und SORE bevorzugt kombinieren [17]. Darüber hinaus gibt es weitere Ansätze zum RE. Etwa dient das Aufgaben-orientierte RE (TORE) nicht als methodisch, festgelegte Vorgehensweise, sondern als mentales Modell. Dieses bietet eine konzeptionelle Rahmenstruktur zum Treffen von Entscheidungen

¹ https://www.ireb.org/content/downloads/1-cpreglossary/ireb_cpre_glossary_17.pdf. [Zuletzt geöffnet am 11.03.2020]

bezüglich der Spezifikation eines Systems unter der Berücksichtigung der richtigen Projektbeteiligten [27].

Des Weiteren existieren Ansätze, die dem RE ähnlich sind und diese ergänzen können, wie etwa die Customer Journey, das Design Thinking [18] oder das User Centered Design (UCD) [1]. Die Customer Journey dient in der Produktentwicklung als Vorgehensmodell, um die mentale Reise eines Kunden zu einer Kaufentscheidung, ausgelöst durch Nutzerbedürfnisse zu beschreiben und verstehen². Design Thinking unterstützt die Entwicklung eines Prototyps als iterative Kreativ-Technik zur Analyse von Problemen und zur Erfüllung von Kundenbedürfnissen. Es basiert auf den aufeinander aufbauenden Arbeitsschritten Problemdefinition, Kundenbedürfnisanalyse, Ideengenerierung durch Brainstorming, Prototyping und Testen [19]. Design Thinking eignet sich für die Kombination mit dem RE, um die Nutzerbedürfnisse zu verstehen und für die Entwicklung eines Prototyps [18]. Das UCD zielt darauf ab, Nutzer frühestmöglich in den Design-Prozess eines Systems zu integrieren. Die ISO 9241-210 [20] definiert einen iterativen UCD-Prozess zur Entwicklung einer Gestaltungslösung, welche die Nutzerbedürfnisse erfüllt. Diese enthält die Aktivitäten Design-Prozess planen, Verstehen und Beschreiben des Nutzungskontexts, Spezifizierung der Nutzungsanforderungen, Entwicklung der Gestaltungslösung gemäß Nutzungsanforderungen und die Gestaltungslösung aus der Nutzerperspektive zu prüfen [20].

Nach mehreren Autoren [21, 22, 23] sind spezifische RE Ansätze für die Entwicklung bestimmter Systeme notwendig. Geis, der Begründer der Kontextmethode, die auf dem UCD basiert, empfiehlt diese besonders für die qualitative Entwicklung von Medizinprodukten und medizinischer Software [21].

2.3 RE-Prozess

Nach Ebert [2] wird der RE-Prozess in die sechs Hauptaktivitäten Ermitteln, Analysieren, Abstimmen, Dokumentieren, Prüfen und Verwalten von Anforderungen unterteilt. Für die Durchführung des RE-Prozesses ist der Anforderungs-Ingenieur verantwortlich, damit kundenorientierte, technische und wirtschaftliche Ziele erreicht werden können [2]. Der Anforderungs-Ingenieur agiert als Bindeglied zwischen den Stakeholdern und allen Beteiligten des RE-Prozesses [2].

Im Folgenden werden die sechs Aktivitäten und Empfehlungen zum RE näher beschrieben. Bei der Anforderungsermittlung werden der Kontext, die Randbedingungen, die Ziele und der Nutzen, des zu entwickelnden Produkts festgelegt. Zusammen mit den ermittelten Stakeholdern können unterschiedliche Techniken zur Anforderungserhebung, wie Befragung oder Beobachtung durchgeführt werden. Durch eine anschließende Auswertung der Ergebnisse werden Anforderungen dokumentiert [24].

Zur Anforderungsanalyse werden die ermittelten Anforderungen durch Analysetechniken und ergänzende Techniken zur Spezifizierung auf ihre Vollständigkeit überprüft. Zu den Analysetechniken zählt etwa die Durchführung von Workshops, die Erstellung einer formellen Anforderungsspezifikation (engl. software requirements specification, SRS) und die Erstellung eines Prototyps zur Annäherung an das Entwicklungsziel.

Die Spezifikation oder Dokumentation von Anforderungen kann durch natürliche Sprache, die Verwendung einer spezifischen Notation, durch zielorientierte Sprachen oder durch die Erstellung von prototypischen Benutzeroberflächen erfolgen [24].

Die Abstimmung von Anforderungen beruht auf dem Konsens der Beteiligten und wird

<https://www.chimpify.de/marketing/customer-journey/>. [Zuletzt geöffnet am 11.03.2020]

durch einen definierten Projektplan mit Vereinbarungen, Arbeitspaketen, Meilensteinen und den erforderlichen Ressourcen bestimmt. Das Verifizieren und Validieren von Anforderungen ist notwendig, um sich ändernde oder neue Anforderungen zu identifizieren und dokumentieren. Es dient zur Überprüfung der Anforderungen auf Richtig- und Vollständigkeit, sowie zur Priorisierung oder Streichung von Anforderungen [2].

Als Prüftechnik eignet sich das Review der SRS oder eines Prototyps zusammen mit den Stakeholdern. Ebenfalls eignen sich Checklisten, die Prüfkriterien zur Erfüllung der Systemziele, der Systemfunktionen und der Gebrauchstauglichkeit des Systems beinhalten [25].

Vollständiges RE basiert auf stetiger Dokumentation, etwa anhand von Entwurfs-Vorlagen und dem Anlegen eines Glossars. Die nachvollziehbare Verwaltung von Anforderungen und ein gezieltes Änderungsmanagement von Anforderungen ist in allen Projektphasen notwendig. Zu den Verwaltungstechniken zählen die Versionierung und eindeutige Identifikation [2] von Anforderungen für die Nachverfolgbarkeit dieser [24].

2.4 RE mit hoher Qualität

Die Durchführung von RE mit hoher Qualität wird durch international anerkannte Standards und Best Practices beschrieben, die Methodenwissen im Bereich RE und Software Engineering vermitteln. Bspw. existiert für die qualitative Entwicklung von Informationssystemen die Norm ISO 29148 [15]. In dieser sind Qualitätsmerkmale für die Erstellung und Pflege von Anforderungen und einer SRS enthalten [2]. Best Practices entsprechen der Durchführung des RE, orientiert am Stand der Technik, etwa nach dem UCD [1] und können im Haftungsfall vor Gericht herangezogen werden [2]. Ein Anforderungs-Ingenieur sollte über das bereits beschriebene Methodenwissen und über Projektmanagement-Kenntnisse verfügen, um das RE als einen festen Bestandteil

des Software-Entwicklung-Prozesses umzusetzen [28]. Agile Software-Entwicklung mit agilem Projektmanagement nach Scrum [29] gewinnt zunehmend an Beliebtheit. Scrum basiert auf kurzen, iterativen Entwicklungszyklen [29] und hat das Ziel nach jedem Entwicklungszyklus ein funktionsfähiges Ergebnis ausliefern zu können [8]. Bei Scrum werden Anforderungen von einem Product Owner als User Storys beschrieben und priorisiert. Diese werden mit dem Entwicklungsteam gemeinsam besprochen und verfeinert. Qualitative User Storys können unter der Berücksichtigung des INVEST-Prinzips [29] erstellt werden. Es dient zur Überprüfung der Akzeptanz, Unabhängig-, Verhandelbar-, Wertig-, Schätzbar-, und Testbarkeit einer User Story. Für die qualitative Umsetzbarkeit und Abnahme von User Storys, können durch das SMART-Prinzip, Akzeptanzkriterien bestimmt werden. Das Akronym SMART steht für die Akzeptanz, Spezifizierbar-, Messbar-, Realisierbar- und Terminierbarkeit eines Akzeptanzkriteriums [29]. Der Einsatz von Scrum sollte die Berücksichtigung von Standards nicht ersetzen. Scrum dient lediglich zum Projektmanagement, das Vorgehen zum RE, der Umgang mit Risiken oder Testfällen wird nicht spezifiziert [29].

Neben den bereits beschriebenen Methodenwissen und den Projektmanagement-Kenntnissen sollte ein Anforderungs-Ingenieur über Fachwissen zum Nutzungskontext eines Systems verfügen [29]. Etwa ist für die Entwicklung von KISs ein breites Verständnis für medizinische Terminologie, als auch die vielseitigen, administrativen, medizinischen und pflegerischen Prozesse notwendig. Außerdem müssen geltende Normen und Gesetze eines bestimmten Landes bzw. einer bestimmten Region berücksichtigt werden [25].

Zusätzlich sollte ein Anforderungs-Ingenieur Fähigkeiten, wie analytisches Denken, Empathie, Kommunikations-, Konflikt- Moderations-, und Überzeugungsfähigkeit als auch Selbstbewusstsein haben [30]. Qualitatives RE beinhaltet zudem die Berücksichtigung von

Anforderungsmanagementtechniken wie Änderungsmanagement, als auch Verhandlungstechniken wie Win-Win-Verhandlungen [24].

3 RE für KISS

In diesem Kapitel wird das RE für Informationssysteme und KISS untersucht, um eine Vergleichsgrundlage des derzeitigen Stands des RE für diese Arbeit zu schaffen, um nachfolgend die Reife des RE des Referenz-KIS-Herstellers bestimmen zu können.

3.1 Studie Entwicklung von Informationssystemen

Folgend werden die Ergebnisse einer Studie zur Entwicklung von Informationssystemen in der Schweiz kurz vorgestellt.

El Emam et al. [24] haben 2012 für die Studie zur Entwicklung von Informationssystemen eine Online-Umfrage zur Messung des Erfolgs von RE durchgeführt. Ziel der Studie ist die Ermittlung von relevanten RE-Zielen und des Zusammenhangs dieser mit erfolgreichem RE. Die Online-Umfrage beinhaltet 419 gültigen Teilnahmen, davon 34 aus dem Gesundheitswesen. Das Ergebnis der Studie beschreibt, dass der Erfolg des RE durch die Prüfung der Bedingungen des RE und der erzielten Qualität eines Produktes gemessen werden kann. Die zu erreichenden Bedingungen des RE ergeben sich aus den Geschäftszielen, Bedürfnissen und Erwartungen der Stakeholder und werden durch den Anforderungs-Ingenieur definiert. Die Bedingungen des RE sind durch die gemeinsame Abstimmung zu dem erzielten Verhalten und den Ausgaben eines Systems, mit den Stakeholdern messbar. Die Qualität des erzielten Produkts basiert auf den werthaltigen und nachverfolgbaren Arbeitsergebnissen des Anforderungs-Ingenieurs. Generell kommt die Studie zu dem Ergebnis, dass das RE weit verbreitet ist, aber nicht einheitlich betrieben wird. Dominante Faktor zur Beurteilung des RE-Erfolgs ergeben sich nicht. Als wichtigste Ziele des RE werden das gemeinsame Verständnis zwischen den Beteiligten am RE-Pro-

zess und eine SRS mit hoher Qualität angegeben. Als RE-Techniken, die besonders in erfolgreichen Entwicklungsprojekten eingesetzt werden, werden Folgende genannt: Workshops zur Ermittlung von Anforderungen, textuelle Szenarios sowie die Beschreibung und Modellierung von Anwendungsfällen zur Analyse und Dokumentation von Anforderungen, die Inspektion zum Prüfen von Anforderungen, das Priorisieren von Anforderungen zum Managen von Anforderungen und das Verhandeln über Anforderungen, etwa durch die Diskussion und Prüfung über einen Umsetzungsvorschlag [24].

3.2 Umfrage Referenz-KIS

Das Referenz-KIS basiert auf einer Software-Lösung für das Patienten- und Fallmanagement für Schweizer Hospitäler. Diese Basis kann durch die Integration von weiteren standardisierten Produktmodulen oder kundenindividuelle Lösungen erweitert werden. Zur Ermittlung des derzeitigen Stands des RE des Referenz-KIS-Herstellers wurden, für diese Arbeit, eine Online-Umfrage durchgeführt. 13 Mitarbeiter des Referenz-KIS-Herstellers, die am RE-Prozess beteiligt sind wurden zu Umfrage eingeladen. Die Rücklaufquote der Umfrage beträgt 77 %.

Zur Herleitung des Fragebogens der Umfrage wurde der Goal-Question-Metrik (GQM) Ansatz angewendet. GQM [31] dient zur Definition von Metriken bei der systematischen Erfassung und Auswertung von Daten. Der Ansatz basiert auf drei Ebenen, die nach dem Top-Down-Prinzip aufeinander aufbauen. Auf der konzeptionellen Ebene wird zuerst das Ziel der Umfrage **F1** formuliert, wie in Tabelle 1 aufgezeigt. Aus **F1** werden auf der operativen Ebene Fragen zur Prüfung der Erreichbarkeit des Ziels abgeleitet und jeweils verfeinert. Zur Beantwortung jeder dieser Fragen werden auf der quantitativen Ebene Metriken bestimmt, wie in Tabelle 1 dargestellt. Der Fragebogen wurde zusätzlich durch die Durchführung eines Pretests optimiert.

Ziel	Frage	Metrik
F1	Welche Ansätze und Aktivitäten umfasst derzeit das RE bei KIS-Herstellern?	n/a
F1.1	Wer führt das RE für die Entwicklung von KIS-Modulen bei KIS-Hersteller mit welcher Qualifikation durch?	M 1.1.1 Menge der Rollen, M 1.1.2 Anteil der Ausbildungsqualifikationen, M1.1.3 Anzahl der Projekterfahrung
F1.2	Für welche Aufgabe in der Software Entwicklung wird das RE bei KIS-Herstellern eingesetzt?	M1.2 Menge und Anteil der Software-Entwicklung
F1.3	Welche Ausgangskriterien sind für das RE bei KIS-Herstellern relevant?	M1.3 Verbreitungsgrad der Ausgangskriterien
F1.4	Was verstehen KIS-Hersteller unter einer Anforderung und wie viele erheben diese in einem Entwicklungsprojekt?	M1.4.1 Verbreitungsgrad der Definition Anforderung, M1.4.2 durchschnittliche Anzahl der Anforderungen in einem Entwicklungsprojekt
F1.5	Welche Vorgehensweisen zum RE gibt es?	M1.5.1 Verbreitungsgrad von Best Practice und Notationen, M1.5.2 Verbreitungsgrad von agilen Methoden
F1.6	Welche Aktivitäten des RE führen KIS-Hersteller in Entwicklungsprojekten durch?	M1.6 Verbreitungsgrad der RE-Aktivitäten
F1.7	Welche Anforderungs-Erhebungstechniken setzen KIS-Hersteller ein?	M1.7 Verbreitungsgrad der Erhebungstechniken
F1.8	Welche Anforderungs-Analysetechniken setzen KIS-Hersteller ein?	M1.8 Verbreitungsgrad der Analysetechniken
F1.9	In welcher Form dokumentieren KIS-Hersteller Anforderungen?	M1.9 Verbreitungsgrad von Dokumentationsformen
F1.10	Wie überprüfen KIS-Hersteller ihre Anforderungen?	M1.10 Verbreitungsgrad von Prüftechniken
F1.11	Welche Verhandlungstechniken nutzen KIS-Hersteller im Rahmen des RE?	M1.11 Verbreitungsgrad von Verhandlungstechniken
F1.12	Welche Kriterien berücksichtigen KIS-Hersteller beim Verwalten von Anforderungen in einem Entwicklungsprojekt?	M1.12 Verbreitungsgrad von Kriterien zur Verwaltung von Anforderungen

Tabelle 1: Verfeinerung der Forschungsfrage 1 nach GQM

3.3 Umfrage Vergleichsgrundlage

In Anlehnung an die Umfrage von El Emam et al. wurde für diese Arbeit eine Online-Umfrage zum RE bei MISs durchgeführt. Der Online-Fragebogen basiert, auf dem gleichen Fragebogen, wie die Umfrage beim Referenz-KIS-Hersteller. Die Datenerhebung erfolgte vom 14. Januar 2020 bis zum 01. Februar 2020. Die Umfrage wurde an 75 Personen in Deutschland, Österreich und der Schweiz verschickt, 50 % davon gehören einem KIS-Hersteller an. Die Rücklaufquote beträgt etwa 20 %, zur Auswertung wurden Fragebögen mit einer Antwortrate von mindestens 50 % berücksichtigt. Die Umfrage richtet sich neben KIS-Herstellern an die Hersteller von MISs, um die Teilnahme und Auswertungsrate der

Umfrage zu erhöhen. Zu den eingeladenen Herstellern von MISs zählen Arztpraxis-, Laborinformationssysteme sowie klinische Bildarchivierungs- und Kommunikationssysteme. Die ausgewerteten Umfrage-Ergebnisse sind in dieser Arbeit anonymisiert und pseudonymisiert dargestellt. Im Folgenden dieser Arbeit werden die Teilnehmer der eben beschriebenen Umfrage als Vergleichsgrundlage bezeichnet.

4 Ergebnisse

In diesem Abschnitt werden die Ergebnisse, der beiden zuvor beschriebenen Umfragen zum Referenz-KIS-Hersteller und der Vergleichsgrundlage in Abbildung 1 präsentiert. Zu jeder Frage aus Tabelle 1 wird die genannte Metrik, in Form von Diagrammen dargestellt. Die Ergebnisse der Antworten zu den 20 gestellten Fragen werden jeweils zu 16 Metriken zusammengefasst. Zum Vergleich der Ergebnisse werden die erhobenen Daten zum Referenz-KIS-Hersteller in Blau und zur Vergleichsgrundlage in Orange dargestellt. Die Antwortrate, wird zu jeder Metrik durch n angegeben. Die Standardabweichung, wird durch SD, aufgezeigt.

Die erhobenen Daten zu M1.1.1 zeigen, dass das RE zur Entwicklung von KISs von mehreren verschiedenen Rollen durchgeführt wird. Eine Person kann mehreren Rollen angehören. Die angegebenen Rollen des Referenz-KIS-Herstellers stimmen größtenteils mit den angegebenen Rollen der Vergleichsgrundlage überein. Beim Referenz-KIS-Hersteller ist zusätzlich der Support am RE-Prozess beteiligt. Bei der Vergleichsgrundlage können Rollen identifiziert werden, die beim Referenz-KIS-Hersteller nicht existieren, wie der Strategist, der Regulatory Affairs Manager und der Usability-Engineer. Generell sind der Anforderungs-Ingenieur und der Produktmanager die verbreitetsten Rollen. Gemäß den ausgewerteten Daten zu M1.1.2 und M1.1.3 variiert die Ausbildungsqualifikation und Projekterfahrung der befragten Personen.

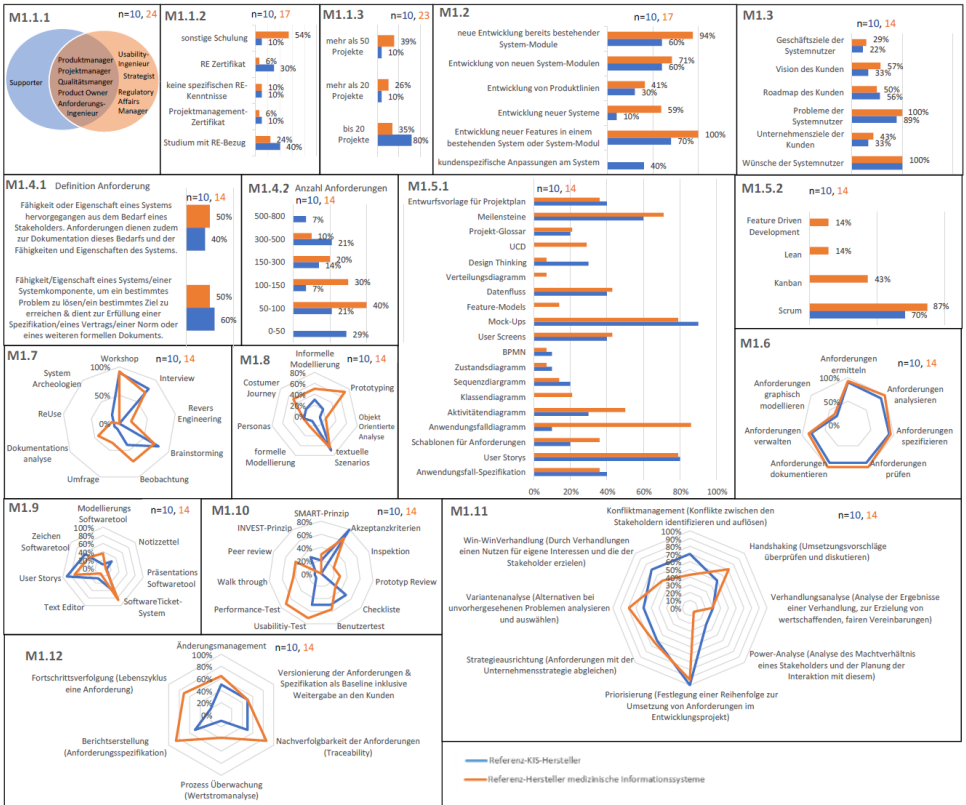


Abbildung 1: Ergebnisse der Umfragen gemäß der Metriken aus Tabelle 1

Das RE wird von den Befragten für die in M1.2 gezeigten Aufgaben eingesetzt. Es ist zu erkennen, dass das RE bei allen Befragten am häufigsten für die Entwicklung von neuen Features, für ein bereits bestehendes System oder Modul, für die Entwicklung von bereits bestehenden und neuen System-Modulen eingesetzt wird. Der Einsatz des RE unterscheidet sich beim Referenz-KIS-Hersteller zur Vergleichsgrundlage bzgl. der Entwicklung von kundenspezifischen Anpassungen. Diese werden lediglich vom Referenz-KIS-Hersteller als Entwicklungsaufgabe angegeben. Für das RE sind für alle Befragten nach M1.3 übereinstimmend besonders die Probleme und Wünsche der Nutzer relevant. Etwa die Hälfte aller Befragten betrachtet bei RE die Roadmap

der Kunden. Die Unternehmens- und Geschäftsziele sowie die Vision der Kunden werden vom Referenz-KIS-Hersteller gegenüber der Vergleichsgrundlage eher berücksichtigt. Die Ergebnisse der Umfragen verdeutlichen zu M1.4.1, dass die Befragten als Anforderung am häufigsten folgende zwei Definitionen verstehen. Von den Befragten definieren 50 % der Vergleichsgrundlage und 40 % des Referenz-KIS-Herstellers eine Anforderung als Fähigkeit oder Eigenschaft eines Systems oder einer Systemkomponente, um ein bestimmtes Problem zu lösen oder ein bestimmtes Ziel zu erreichen. Darüber hinaus dient eine Anforderung zur Erfüllung einer Spezifikation, eines Vertrags, einer Norm oder eines weiteren formel-

len Dokuments [10]. Von den Befragten verstehen außerdem 60 % des Referenz-KIS-Herstellers und 40 % des Referenz-KIS-Herstellers unter einer Anforderung: Die Fähigkeit oder Eigenschaft eines Systems hervorgegangen aus dem Bedarf eines Stakeholders. Anforderungen dienen zudem zur Dokumentation dieses Bedarfs und der Fähigkeiten und Eigenschaften des Systems. Nach den erhobenen Daten zu M1.4.2 erhebt der Referenz-KIS-Hersteller in einem Entwicklungsprojekt durchschnittlich 125 Anforderungen (SD=229) und die Vergleichsgrundlage 150 Anforderungen (SD=119).

Zur Unterstützung des RE-Prozesses werden von allen Befragten nach den Ergebnissen zu M1.5.1 besonders Mock-Ups, User Storys und Meilensteine als Hilfsmittel eingesetzt. Mehr als 80 % der Befragten der Vergleichsgrundlage Erstellen im RE-Prozess Anwendungsfalldiagramme. Bei Referenz-KIS-Hersteller werden hingegen von 10 % der Befragten Anwendungsfalldiagramme einsetzen. Der Einsatz von speziellen Entwurfsvorlagen und Modellen zur Orientierung an Best Practices erfolgt bei allen Befragten geringfügig. Angaben zur Berücksichtigung von Standards wurden von den Befragten nicht gemacht. Teilweise werden nach den Angaben aller Befragten Datenfluss-Diagramme, User-Screens, Aktivitäts-Diagramme, Anwendungsfall-Spezifikationen und Entwurfsvorlagen Projektpläne als Hilfsmittel genutzt. Für die Durchführung des agilen Projektmanagements wird nach M1.5.2 am häufigsten Scrum verwendet. Nach den Angaben der Befragten der Vergleichsgrundlage werden neben Scrum noch weitere Rahmenwerke zum agilen Projektmanagement eingesetzt.

Gemäß M1.6 werden die Aktivitäten des RE-Prozesses zum Ermitteln, Spezifizieren, Analysieren, Prüfen, Dokumentieren und Verwalten von Anforderungen größtenteils von allen Befragten durchgeführt. Anforderungen werden von weniger als der Hälfte der Befragten graphisch modelliert.

Nach der Auswertung zu M1.7 gehören zu den häufigsten Erhebungstechniken von Anforderungen, die Durchführung von Workshops, Brainstorming, Interviews, und Beobachtungen. Im Vergleich zur Vergleichsgrundlage führt der Referenz-KIS-Hersteller weniger Beobachtungen zur Anforderungserhebung durch, stattdessen sind die Angaben zur Durchführung von Interviews und Brainstormings beim Referenz-KIS-Hersteller höher.

Die drei häufigsten durchgeführten Techniken zur Analyse von Anforderungen sind nach den ausgewerteten Angaben zu M1.8 textuelle Szenarios, Prototyping, informelle Modellierung und Customer Journey. Die Angaben des Referenz-KIS-Herstellers übersteigen die der Vergleichsgrundlage, bezogen auf die Erstellung von textuellen Szenarios. Hingegen ist der Anteil des Referenz-KIS-Herstellers zu den weiteren Analysetechniken geringer.

Die Datenerhebung zu M1.9 verdeutlicht, dass der Großteil aller Befragten Anforderungen in einem Ticketsystem dokumentiert. Etwa 75 % davon dokumentieren Anforderungen als User Storys.

Passend dazu wird nach M1.10 von allen Befragten die Überprüfung von Anforderungen nach Akzeptanzkriterien als eine der häufigsten Prüfetechniken angegeben. Zusätzlich werden von der Vergleichsgrundlage, im Gegensatz zum Referenz-KIS-Hersteller häufiger Performance-, Usability- und Benutzertest als Prüfetechnik durchgeführt. Der Einsatz von Checklisten zur Überprüfung von Anforderungen erfolgt beim Referenz-KIS-Hersteller häufiger.

Nach der Auswertung der Daten zu M1.11 dient die Priorisierung von Anforderungen als meist genutzte Verhandlungstechnik zum Managen von Anforderungen. Handshaking und die Variantenanalyse werden von der Vergleichsgrundlage häufiger angewendet als vom Referenz-KIS-Hersteller.

Gemäß den erhobenen Daten zu M1.12 sind nach allen Befragten die Erstellung einer SRS, die Nachverfolgbarkeit und Änderungsmanagement von Anforderungen die wichtigsten

Aspekte zur Verwaltung von Anforderungen. Bis auf die Tätigkeit der Versionierung von Anforderungen werden alle in M1.12 dargestellten Anforderungs-Verwaltungstechniken von der Vergleichsgrundlage häufiger ausgeführt als vom Referenz-KIS-Hersteller.

5 Diskussion

Zur Analyse des RE im Bereich KISs basiert diese Arbeit auf der Ermittlung des Stands der Technik zu MISs und auf der Durchführung von zwei Experten-Umfragen für diese Arbeit. Die Ergebnisse dieser Arbeit zeigen, dass die befragten Experten der Hersteller von KIS und MIS die Aktivitäten des RE-Prozesses nach Ebert [2] größtenteils berücksichtigen und die Rolle des Anforderungs-Ingenieures etabliert ist. Durch die für diese Arbeit ermittelten Metriken und Ergebnisse der beiden Umfragen konnte der Verbreitungsgrad von RE-Techniken, für die Software-Entwicklung von KISs aufgezeigt werden. Diese Arbeit zeigt, dass zur Ermittlung von Anforderungen am häufigsten die Durchführung von Workshops und Brainstormings vorgenommen wird. Die meist verwendeten Techniken zur Analyse von Anforderungen sind die Erstellung von textuellen Szenarios und Prototypen. Akzeptanzkriterien dienen als meist genutzte Prüftechnik und User Storys als die häufigste Dokumentationsform von Anforderungen, da der Großteil der Hersteller agiles Projektmanagement betreibt. Die Nachverfolgbarkeit und Priorisierung von Anforderungen dient als verbreitetste Technik zum Managen von Anforderungen. Die Erstellung einer SRS wird als wichtigste Voraussetzung für die Verhandlung mit Kunden angesehen. Die erhobenen Ergebnisse dieser Arbeit stimmen mit den Ergebnissen der Studie von El Emam et al. [24] bzgl. des Verbreitungsgrads der RE-Techniken Workshops, textuelle Szenarios und Priorisierung von Anforderungen überein. Der Vergleich der Ergebnisse der Studie von El Emam et.al und der Ergebnisse dieser Arbeit zeigt, dass das RE nicht einheitlich praktiziert wird, aber weit verbreitet ist. Einigkeit besteht zur Studie von El Eman et.

al, sowie zu Forschung von Valentini [6], dass für ein qualitatives RE, ein gemeinsames Verständnis zum RE-Prozess und der Beschaffenheit der Anforderungen unter den Beteiligten bestehen muss. Diese Arbeit kommt zudem zum Ergebnis, dass die Ausbildung, die Fähigkeiten, der Wissens- und der Erfahrungsschatz des Anforderungs-Ingenieures, die Reife des RE bei KISs beeinflussen. Zur Ermittlung der Reife des RE des Referenz-KIS-Herstellers wurden in diese Arbeit Übereinstimmungen und Abweichungen zum RE des Referenz-KIS-Herstellers gegenüber der Vergleichsgrundlage sowie der Empfehlungen der Literatur identifiziert. Die Ursachen und Auswirkungen dieser Abweichungen, wurden in dieser Arbeit nicht analysiert.

Die Ergebnisse der Arbeit beziehen sich lediglich auf die Angaben des Referenz-Herstellers im Vergleich zu den durchschnittlichen Angaben der Vergleichsgrundlage. Die Rücklaufquote des Referenz-KIS-Herstellers und der Vergleichsgrundlage sind nicht gleichwertig. Es erfolgt zudem keine differenzierte Betrachtung zu den Angaben der einzelnen Mitarbeiter des Referenz-KIS-Herstellers oder der verschiedenen Hersteller von MISs. Hierfür sind weitere Analysen zu den erhobenen Daten notwendig. Des Weiteren wird in dieser Arbeit der Einsatz des RE nicht auf eine speziell bezogene Aufgabe der Software-Entwicklung für KISs untersucht. Die Durchführung von weiteren Untersuchungen zum RE zu den einzelnen Aufgaben der Software-Entwicklung für KISs würde diese Forschung ergänzen.

6 Fazit und Ausblick

In dieser Arbeit wurde gezeigt, dass RE ein wichtiger Bestandteil für die KIS-Entwicklung ist und auf einem gemeinsamen Verständnis aller Beteiligten beruht. Die Arbeit hat verschiedene Vorgehensweisen, Aktivitäten, Techniken, Best Practice und Standards zum RE vorgestellt. Dem Referenz-KIS-Hersteller wird, basierend auf dieser Arbeit, die Berücksichtigung dieser ausgewählten Best Practices des RE für die zukünftige KIS-Entwicklung

empfohlen. Zudem wird dem Referenz-KIS-Hersteller die Auswahl von geeignetem RE nach der Aufgabe der Software-Entwicklung empfohlen. Diese Arbeit soll für eine weiteren Forschung, im Rahmen einer Masterarbeit, zur Konzeptionierung einer neuer RE-Methode für die KIS-Entwicklung dienen. Die neue Methode soll die Vorteile verschiedener Ansätze und Best Practices des RE kombiniert, um eine Auswahl des RE bezogen auf die Aufgabe der KIS-Entwicklung zu ermöglichen.

7 Literatur

- [1] L. Teixeira, C. Ferreira und B. S. Santos, User-centered requirements engineering in health information systems: A study in the hemophilia field, *Computer Methods and Programs in Biomedicine*, Jg. 106, Nr. 3, S. 160–174, 2012, ISSN: 0169-2607.
- [2] C. Ebert, *Systematisches Requirements Engineering: Anforderungen ermitteln, dokumentieren, analysieren und verwalten*, 6th ed. Heidelberg: dpunkt. Verl., 2019, ISBN: 9783864905629.
- [3] P. Haas, *Medizinische Informationssysteme und elektronische Krankenakten*. Berlin: Springer, 2005, ISBN: 9783540204251.
- [4] D. Dermeval, J. Vilela, I. I. Bittencourt, J. Castro, S. Isotani, P. Brito und A. Silva, Applications of ontologies in requirements engineering: a systematic review of the literature, *Requirements Engineering*, Jg. 21, Nr. 4, S. 405–437, 2016, ISSN: 0947-3602.
- [5] J. Horkoff, F. B. Aydemir, E. Cardoso, T. Li, A. Maté, E. Paja, M. Salnitri, L. Piras, J. Mylopoulos und P. Giorgini, Goal oriented requirements engineering: an extended systematic mapping study, *Requirements Engineering*, Jg. 24, Nr. 2, S. 133–160, 2019, DOI: 10.1007/s00766-017-0280-z.
- [6] U. Valentini, R. Weißbach, R. Fahney, T. Gartung, J. Glunde, A. Herrmann, A. Hoffmann und E. Knauss, *Requirements Engineering und Projektmanagement*, Ser. Xpert.press. Berlin: Springer, 2013, ISBN: 978-3642294310.
- [7] J. Patton, P. Economy und P. Hildebrandt, *User story mapping: Die Technik für besseres Nutzerverständnis in der agilen Produktentwicklung*, Dt. Ausg., 1th ed., Beijing: O'Reilly, 2015, ISBN: 9783958750678.
- [8] I. Sommerville, *Software Engineering*, 10th ed, Ser. it - informatik. Hallbergmoos: Pearson, 2018, ISBN: 9783863268350.
- [9] Ludewig; Jochen (2014): *Software-Prozesse und Software-Qualität*. In: *Wechselwirkungen, Jahrbuch aus Lehre und Forschung der Universität Stuttgart* 2006, S. 52-59.
- [10] ISO/IEC/IEEE 24765:2010, International Organization for Standardization, *Systems and software engineering - Vocabulary*
- [11] ISO/IEC 15288:2008, International Organization for Standardization, *Systems and software engineering -System life cycle processes*
- [12] R. Wieringa, Requirements researchers: are we really Doing research?, *Requirements Engineering*, Jg. 10, Nr. 4, S. 304–306, 2005, DOI: 10.1007/s00766-005-0013-6.
- [13] K. Decreus, M. Snoeck und G. Poels, Practical Challenges for Methods Transforming i* Goal Models into Business Process Models, in *17th IEEE 2009 International Requirements Engineering Conference*, IEEE, S. 15–23, ISBN: 978-0-7695-3761-0.
- [14] A. van Lamsweerde, Goal-oriented requirements engineering: a guided tour, in *Proceedings Fifth IEEE International Symposium on Requirements Engineering*, IEEE Comput. Soc, 2001, S. 249–262, ISBN: 0-7695-1125-2.
- [15] ISO/IEC/IEEE 29148, International Organization for Standardization, *Systems and software engineering - Life cycle processes - Requirements engineering*, 2008

- [16] Robertson, S.; Robertson, J., Reliable requirements through the quality gateway. In: Proceedings. 10th International Workshop on Database and Expert Systems Applications. DEXA 99. Florence, Italien, 1999: IEEE, S. 357–363, DOI:10.1109/DEXA.1999.795193
- [17] E. Sikora, Ein modellbasierter Ansatz zur verzahnten Entwicklung von Anforderungen und Architektur über mehrere Abstraktionsstufen hinweg: Zugl.: Duisburg-Essen, Univ., Diss., 2010. Berlin: Logos-Verl., 2011, ISBN: 9783832527624.
- [18] N. Carroll und I. Richardson, Aligning healthcare innovation and software requirements through design thinking, in Proceedings of the International Workshop on Software Engineering in Healthcare Systems – SEHS '16, New York, USA: ACM Press, 2016, S. 1–7, ISBN: 9781450341684.
- [19] C. P. Hoffmann, S. Lennerts, C. Schmitz, W. Stölzle und F. Uebernickel, Business Innovation: Das St. Galler Modell. Springer Wiesbaden, 2016. DOI: 10.1007/978-3-658-07167-7.
- [20] DIN EN ISO 9241-210, DIN e. V., Ergonomie der Mensch System-Interaktion - Teil 210: Prozess zur Gestaltung gebrauchstauglicher interaktiver Systeme, 2011.
- [21] T. Geis und C. Johnner, Hrsg., Usability Engineering als Erfolgsfaktor: Effizient IEC 62366- und FDA-konform dokumentieren, 1.th ed., Ser. Praxis. Berlin: Beuth, 2015, ISBN: 9783410245001.
- [22] G. A. Lewis, E. Morris, P. Place, S. Simanta und D. B. Smith, Requirements engineering for systems of systems, in 3rd Annual IEEE Systems Conference, 2009, Piscataway, NJ: IEEE, S. 247–252, ISBN: 978-1-4244-3462-6.
- [23] P. Clements und L. Northrop, Software product lines: Practices and patterns, 7. print, Ser. SEI series in software engineering. Boston u. a.: Addison-Wesley, 2009, ISBN: 0134424085.
- [24] S. A. Fricker, R. Grau und A. Zwingli, Requirements Engineering: Best Practice, in Requirements Engineering for Digital Health, S. A. Fricker, C. Thümmel und A. Gavras, Hrsg., Bd. 18, Springer, 2015, S. 25–46, ISBN: 978-3-319-09797-8.
- [25] S. A. Fricker, Requirements engineering for digital health. Springer, 2015, ISBN: 978-3-319-09798-5.
- [26] W. Abramowicz und A. Kokkinaki, Business Information Systems Workshops. Springer, 2014, Bd. 183, ISBN: 978-3-319 - 11459-0.
- [27] Adam, Sebastian; Doerr, Joerg; Eisenbarth, Michael; Gross, Anne: Using Task-oriented Requirements Engineering in Different Domains - Experiences with Application in Research and Industry. 17th IEEE International Requirements Engineering Conference Atlanta 2009. S. 267–272, DOI: 10.1109/RE.2009.16.
- [28] K. Pohl, G. Böckle und F. Linden, Software Product Line Engineering: Foundations, Principles, and Techniques. Berlin: Springer-Verlag, 2005, ISBN: 978-3-540-24372-4.
- [29] J. Bergsmann und M. Unterauer, Requirements Engineering für die agile Softwareentwicklung: Methoden, Techniken und Strategien zur erfolgreichen Umsetzung, 1th ed., Heidelberg: dpunkt-Verl., 2014, ISBN: 978-3-86490-149-2.
- [30] K. Pohl und C. Rupp, Basiswissen Requirements Engineering: Aus- und Weiterbildung nach IREB-Standard zum Certified Professional for Requirements Engineering: foundation level nach IREB-Standard, 4th ed. Heidelberg: dpunkt, 2015, ISBN: 3864902835.
- [31] Jochen Ludewig und Horst Lichter, Software Engineering: Grundlagen, Menschen, Prozesse, Techniken. 2nd ed, Heidelberg: dpunkt-Verl., 2010, ISBN: 3898646629.

Traceability für Usability Engineering bei medizintechnischen Produkten

Antonia Schöning

Hochschule Reutlingen

Antonia.Schoening@Student.Reutlingen-University.de

Abstract

Die Entwicklung eines Medizinproduktes benötigt in der Regel mehrere Jahre. Gesetzliche Vorgaben, wie zum Beispiel das Medizinprodukte Durchführungsgesetz, bestimmen, welche Schritte während der Entwicklung durchgeführt werden müssen. Deren Einhaltung muss in der technischen Dokumentation nachgewiesen werden. Die darin enthaltenen technischen Dokumente entstehen im Verlauf der Entwicklung. Diese bauen aufeinander auf und verweisen sich gegenseitig. Dadurch entstehen heterogene und unübersichtliche Strukturen. Eine Lösung für dieses Problem bietet Traceability. Traceability sorgt dafür, dass die Anforderungen an das Medizinprodukt mit Dokumenten, wie dem Anforderungskatalog, Lastenheft oder der Spezifikation verknüpft werden können. Somit ist jederzeit nachvollziehbar, welche Anforderungen mit welchem Test, welchen Änderungen oder welchen Ergebnissen zusammenhängen. Ein wichtiger Prozess bei der Entwicklung von Medizinprodukten ist zudem das Usability Engineering, wodurch die Sicherheit eines Medizinprodukts sichergestellt und Risiken bei der Anwendung

minimiert werden sollen. In diesem Prozess entstehen viele Artefakte, wie zum Beispiel Usability-Berichte. Um den Überblick über alle Usability-Daten behalten zu können, können diese mithilfe von Traceability verknüpft werden. In diesem Artikel wird herausgestellt, welche Voraussetzungen für das Usability Engineering in der Medizintechnik an Traceability gestellt werden.

Schlüsselwörter

Medizintechnik, Traceability, Usability Engineering

CR-Kategorien

- Security and privacy
 - ~ Human and societal aspects of security and privacy
 - ~ Usability in security and privacy
- Information systems
 - ~ Data management systems
 - ~ Data structures

1 Einleitung

Medizintechnische Produkte werden zur Behandlung oder Unterstützung von erkrankten Menschen eingesetzt [1]. Im Extremfall kann die Sicherheit eines Medizinproduktes das Leben eines Menschen gefährden. Diabetespatienten können zum Beispiel eine Pumpe verwenden, die ihnen das benötigte Hormon Insulin regelmäßig zuführt. Eine hohe Sicherheit der Pumpe bedeutet, dass technisches Versagen und die falsche Bedingung durch den Anwender so selten wie möglich auftreten.

Betreuer Hochschule: Prof. Dr. Christian Kücherer
Hochschule Reutlingen
Christian.Kuecherer@Reutlingen-University.de

Wissenschaftliche Vertiefungskonferenz
20. Mai 2020, Hochschule Reutlingen
Copyright 2020 Antonia Schöning

Für ein möglichst geringes Anwendungsrisiko muss das Medizinprodukt technisch fehlerfrei funktionieren und einen hohen Grad an Usability aufweisen [2]. Dieser gibt an, wie anwenderfreundlich das Produkt gestaltet ist. Da die Sicherheit bei einem Medizinprodukt einen hohen Stellenwert hat, sind MedizinproduktHersteller gesetzlich dazu verpflichtet, die Sicherheit ihrer Produkte zu prüfen und zu dokumentieren. Aus der Dokumentation muss hervorgehen, wie die Anforderungen evaluiert worden sind und welche Ergebnisse und Änderungen daraus resultieren [3]. In der Praxis besteht die geforderte Dokumentation aus vielen einzelnen Dokumenten und Artefakten, wie zum Beispiel dem Quellcode einer Software. Aufgrund der großen Menge ist es eine Herausforderung, die Dokumente und Artefakte sinnvoll und übersichtlich miteinander zu verknüpfen. Dies ist jedoch notwendig, um nachvollziehen zu können, ob alle Anforderungen umgesetzt und geprüft worden sind. Eine Lösung für dieses Problem bietet Traceability [4]. Traceability ermöglicht es, Verbindungen und Beziehungen zwischen Dokumenten oder Artefakten herzustellen. Im Usability Engineering wird es bisher selten verwendet, da die meisten Usability-Artefakte erst im späteren Verlauf der Produktentwicklung entstehen. Zu diesem Zeitpunkt haben die Hersteller bereits Tabellen angelegt, die einen Überblick über die Entwicklungsartefakte geben. Da Traceability jedoch die Qualität und Verwaltung der Dokumentation und Usability-Artefakte verbessern kann, werden in diesem Artikel Voraussetzungen ermittelt, die das Usability Engineering aus der Medizintechnik an Traceability stellt [4].

Zunächst werden in Kapitel 2 die Grundlagen erläutert, die zum Verständnis der Thematik notwendig sind. Anschließend wird in Kapitel 3 dargestellt, wie sich das allgemeine Usability Engineering vom Usability Engineering in der Medizintechnik unterscheidet, welche Artefakte dabei entstehen und welche gesetzlichen Vorgaben einzuhalten sind. In Kapitel 4

werden Nutzer von Traceability in der Medizintechnik skizziert, sowie der Nutzen und eine Einsatzmöglichkeit von Traceability dargestellt. Schließlich werden die Ergebnisse in Kapitel 5 zusammengefasst und ein Ausblick auf ein zukünftiges Projekt gegeben.

2 Grundlagen

Dieses Kapitel beinhaltet Grundlagen und Definitionen zum Verständnis der Bereiche Usability Engineering, Entwicklung von Medizinprodukten und Traceability. In Abschnitt 2.1 ist die Bedeutung von Usability und der Prozess des Usability Engineerings erläutert. Abschnitt 2.2 erklärt, was bei der Entwicklung von Medizinprodukten zu beachten ist. Schließlich werden in Abschnitt 2.3 die Begriffe *Trace* und *Traceability* erläutert.

2.1 Usability Engineering

Der Begriff *Usability* kann mit *Gebrauchstauglichkeit* oder *Benutzerfreundlichkeit* übersetzt werden. Die DIN EN ISO 9241-110 [5] definiert Usability als Maß wie effektiv, effizient und zufriedenstellend die Anwender eines Systems in ihrem Nutzerkontext mithilfe des Systems an ihr Ziel gelangen. Usability Engineering ist der Prozess, der durchlaufen wird, um eine möglichst gute Usability zu erzielen. Es gibt verschiedene Usability Engineering Prozesse. Häufig wird das Vorgehen nach der DIN EN 9241-210 [6] gewählt. Die Norm beschreibt ein vierstufiges iteratives Vorgehen, um an ein hohes Maß an Usability zu gelangen. Abbildung 1 stellt den Usability Engineering Prozess gemäß der DIN EN 9241-210 dar. Nach der Planung des iterativen Vorgehens, wird zuerst (1) der Nutzerkontext der Anwender ermittelt. Dazu gehören zum Beispiel die Aufgaben, Abläufe und die Umgebung der Anwender. Aus der Analyse des Nutzerkontextes können (2) Anforderungen ermittelt und spezifiziert werden. Diese werden in (3) einem Systementwurf umgesetzt, welcher (4) gemeinsam mit den Anwendern getestet wird. Die einzelnen Schritte werden iterativ

wiederholt, bis das System alle zuvor definierten Nutzungsanforderungen erfüllt und effektiv, effizient und zufriedenstellend ist [6].

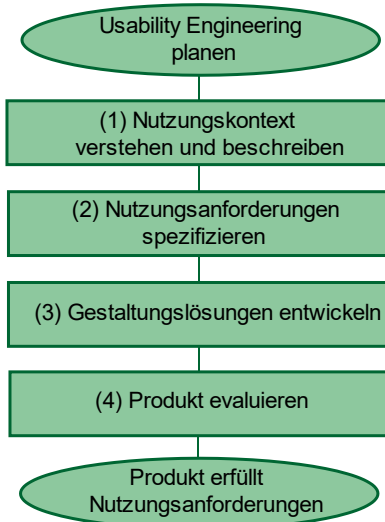


Abbildung 1: Usability Engineering Prozess nach DIN EN 9241-210 [6]

2.2 Systementwicklung in der Medizintechnik

Ein Medizinprodukt ist ein Gegenstand, Softwaresystem oder Stoff, der für diagnostische oder therapeutische Zwecke am Menschen eingesetzt wird [1]. Das Spektrum von Medizinprodukten reicht von Gegenständen die Leben retten können, wie zum Beispiel Implantate, über Geräte im Krankenhaus, wie Ultraschallsysteme, bis zu kleinen Hilfsmitteln, welche die Lebensfreude der Anwender erhöhen. Die Produkte werden in vier Risikoklassen eingeteilt. Je gefährlicher die Anwendung eines Produktes ist, desto höher ist die Risikogruppe. Bei hohen Risikogruppen müssen strengere Abnahmeprüfungen bestanden werden, bevor das Produkt eingesetzt werden kann [7].

Die Entwicklung aller Medizinprodukte lässt sich in fünf Phasen unterteilen, bei der Ingenieure und Mediziner häufig zusammenarbeiten. Zu Beginn steht (1) die Idee eines Produkts, die häufig aus Forschungsprojekten entsteht. Hierauf folgt (2) die Entwurfsphase, in der Prototypen getestet werden bis schließlich das fertige Produkt hergestellt werden kann [7]. Medizinprodukte müssen zertifiziert werden, bevor sie verkauft werden können. Das geschieht in der Phase (3) der Marktzulassung, in der die Einhaltung aller Richtlinien überprüft wird. Die Wirksamkeit, Leistungsfähigkeit und Sicherheit stehen stets im Zentrum der Entwicklung. Um diesen drei Bedingungen nachzukommen, müssen die Hersteller eine Risikoanalyse, eine klinische Bewertung zur Leistungsfähigkeit und umfangreiches Qualitätsmanagement durchführen. In Phase 4 durchlaufen die Produkte in Deutschland ein Verfahren zur Erstattung der Kosten durch die Krankenkassen. Schließlich kann das Produkt in der letzten Phase (5) auf den Markt gebracht werden [2].

2.3 Traceability

Nach Cleland-Huang [4] wird Traceability (dt. Rückverfolgbarkeit) als Fähigkeit verstanden, eine Spur zwischen Artefakten im Entwicklungsprozess herzustellen und zu verwenden. Diese Spuren werden als *Trace* bezeichnet. In der Systementwicklung besteht ein Trace aus drei Komponenten: dem Quellartefakt, dem Zielartefakt und dem Link, also der Verbindung zwischen den beiden Artefakten. Zu den Artefakten zählen unter anderem Anforderungen, Tests und Quellcode einer Software. Das Quellartefakt kann zum Beispiel eine Anforderung sein, deren Link auf einen Testaufbau verweist. Ein Artefakt kann mit mehreren Artefakten verlinkt sein und somit zeitgleich Quell- und Zielartefakt in unterschiedlichen Traces sein. Es gibt verschiedene Arten von Links mit denen die Traces geordnet und Analysen durchgeführt werden können. Zwei Beispiele für Link-Arten sind *wird erfüllt durch* und *wird getestet durch*: „Anforderung A wird erfüllt durch

Code C“ oder „Code C wird getestet durch Test T “. [4]

3 Usability Engineering in der Medizintechnik

Usability ist ein wichtiges Qualitätsmerkmal von Software. Sie sorgt dafür, dass die Anwender des Systems zufrieden an ihr Ziel gelangen. In der Medizintechnik geht der Nutzen von Usability darüber hinaus: Usability Engineering wird eingesetzt, um die Sicherheit in Gefahrensituationen zu erhöhen. In Abschnitt 3.1 wird der Prozess des Usability Engineerings in der Medizintechnik dargestellt. Die Usability-Artefakte, die während des Prozesses entstehen, werden in Abschnitt 3.2 erläutert. Abschnitt 3.3 verdeutlicht die gesetzlichen Grundlagen, denen das Usability Engineering in der EU und den Vereinigten Staaten von Amerika unterliegt.

3.1 Unterschied zum allgemeinem Usability Engineering

Beim Usability Engineering in der Medizintechnik können dieselben Usability-Methoden eingesetzt werden, wie im allgemeinen Usability Engineering. Der Unterschied ist, dass es in der Medizintechnik nicht nur um Effektivität, Effizienz und Zufriedenstellung geht, sondern auch um die Sicherheit vor Gefährdungen während der Anwendung des Medizinprodukts. Laut der IEC 62366-1 [8] soll Usability bei der Entwicklung von Medizinprodukten die Sicherheit bei der Nutzung des Produktes maximieren. Deswegen ist Usability Engineering in der Medizintechnik ein Teil des Risikomanagements. Ein wichtiger Fachbegriff in diesem Zusammenhang ist der *Use Error*. Ein Use Error ist eine Handlung oder die Unterlassung einer Handlung bei der Nutzung des Medizinprodukts, die zu einem anderen Ergebnis führt, als die Anwender es erwarten oder es vom Hersteller vorgesehen wurde. Zum Beispiel entsteht ein Use Error, wenn ein Diabetespatient an seiner Insulinpumpe versehentlich das Zehnfache der

gewünschten Insulinmenge einstellt. Anhand des Usability Engineerings sollen potentielle Use Errors ermittelt, dokumentiert und nach Möglichkeit verhindert werden [8]. Der Use Error an der Insulinpumpe könnte verhindert werden, indem der Patient sehr hohen Angaben zusätzlich bestätigen muss oder das System Eingaben über einem Maximalwert nicht zulässt.

Der allgemeine Prozess des Usability Engineerings besteht aus vier iterativen Schritten, wie in Abbildung 1 dargestellt. In der Medizintechnik eignet sich eine Unterteilung nach der IEC 62366-1 in neun Schritte, die alle dokumentationspflichtig sind. Der Prozess ist in Abbildung 2 abgebildet. Im ersten Schritt muss der Hersteller die *Use Specification* erstellen, welche unter anderem den Zweck des Medizinprodukts beschreibt. Im zweiten und dritten Schritt werden die oben genannten Use Errors und Gefährdungen bei der Verwendung des Produktes ermittelt. Dazu werden im vierten Schritt Situationen, sogenannte *gefährdungsbezogene Use Scenarios*, erstellt, in denen dem Nutzer eine Gefahr bei der Verwendung des Produktes droht. Die Ergebnisse werden im fünften Schritt für die Evaluation priorisiert. Als sechster Schritt muss das Design des Produktes so entwickelt oder überarbeitet werden, dass möglichst wenig der ermittelten Gefahrensituationen entstehen können. Im siebten Schritt werden die formative und die summative Evaluation geplant. Die formative Evaluation dient der Bewertung von Zwischenergebnissen und ermöglicht es Methoden oder Vorgehensweisen im Projekt zu verändern, wenn das Zwischenergebnis nicht zufriedenstellend ist. Diese wird in Schritt acht durchgeführt. Am Ende des Projekts folgt in Schritt neun die summative Evaluation, die das Endprodukt bewertet. [8]

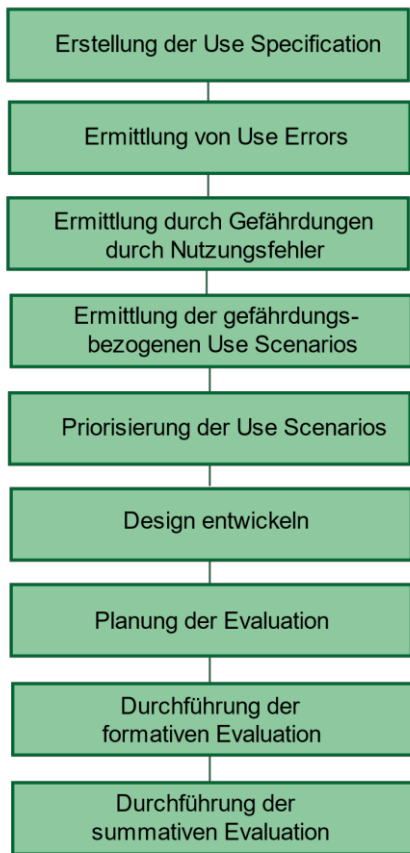


Abbildung 2: Usability Engineering Prozess in der Medizintechnik nach IEC 62366-1 [8]

3.2 Artefakte aus dem Usability Engineering in der Medizintechnik

Beim Usability Engineering hängt die Art der Artefakte von den gewählten Usability-Methoden ab. Bei einem Interview entstehen ein Fragenkatalog, Gesprächsnotizen, Tonaufnahmen und Auswertungen. Diese können in unterschiedlichsten Dateiformaten vorliegen. Der Fragenkatalog kann mit einem Textverarbeitungsprogramm erstellt worden sein, die

Gesprächsnotizen wurden handschriftlich angefertigt und eingescannt, die Tonaufnahme ist in einem Audioformat und die Auswertung ist in einer Tabellenkalkulation zu finden. Der gesamte Prozess des Usability Engineerings muss dokumentiert werden, um den Prozess nachvollziehbar und beweisbar zu gestalten.

Die IEC 26466-1 [8] beinhaltet alle vorgeschriebenen Pflichtartefakte und kann somit als Leitfaden dienen. Das Hauptartefakt ist der *Usability Engineering File*. Darin werden alle Ergebnisse des gesamten Usability Engineering Prozesses festgehalten. Die Ergebnisse müssen nicht physisch enthalten sein, sondern können auch auf andere Artefakte verweisen. Der Usability Engineering File kann Teil von anderen Akten des Entwicklungsprozesses sein, wie zum Beispiel der Risikomanagementakte oder der Produktdesign-Akte. In Abschnitt 3.1 wurde die *Use Specification* bereits erwähnt. Diese ist der erste Teil im Usability Engineering File. Sie gibt den medizinischen Zweck des Produktes an und enthält Informationen zur Patientengruppe und den Nutzungskontext. Im Usability Engineering Prozess werden Gefahren bei der Verwendung des Produktes ermittelt. Diese werden in Form von *gefährdungsbezogenen Use Scenarios* festgehalten. Sie beinhalten die gefährdungsbezogene Situation, die durchzuführenden Aufgaben der Anwender in dieser Situation und den Schweregrad des Schadens, der entstehen könnte. Darauf folgt die *User Interface Specification*. Diese beschreibt die sicherheitsrelevanten Gründe für das Design des Medizinprodukts und muss angeben ob ein Training oder das Lesen der Begleitdokumentation zur Anwendung des Produktes erforderlich ist. Schließlich wird die *User Interface Evaluation* angelegt. Diese besteht aus dem Plan und den Ergebnissen zur Usability-Evaluation des Produkts und enthält die Dokumentation der ausgewählten Usability-Methoden der formativen und summativen Evaluation. Abbildung 3 fasst die vorgeschlagene Ordnung der IEC 26466-1 zusammen.

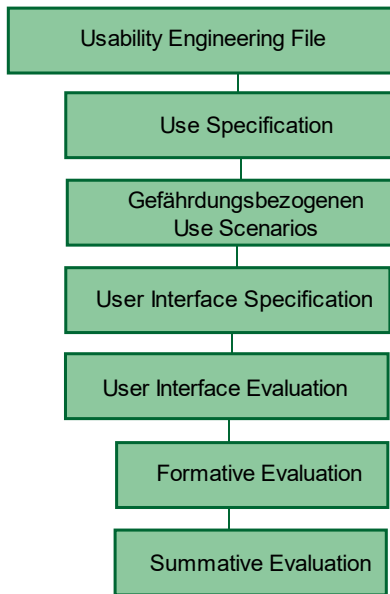


Abbildung 3: Dokumentationsplan des Usability Engineerings nach IEC 62366-1 [8]

3.3 Gesetzliche Grundlagen zum Usability Engineering in der Medizintechnik

Der komplette Ablauf der Entwicklung von Medizinprodukten ist nicht im Detail vorgeschrieben. Jedoch müssen bestimmte Schritte in der Entwicklung eingehalten werden. Welche Schritte das sind, hängt von den gesetzlichen Vorgaben des Landes ab, in dem das Produkt vermarktet werden sollen. Alle Vorgaben ordnet die Dokumentation des Usability Engineerings an.

In Amerika gilt der Code of Federal Regulation (21 CFR) [9], in dem die amerikanische Gesundheitsbehörde Food and Drug Administration (FDA) festlegt, dass die Rückverfolgbarkeit der Entwicklung jedes Medizinprodukts gewährleistet sein muss. Usability wird von der FDA eingesetzt, um

Risiken bei der Anwendung des Produktes zu minimieren. Zur Dokumentation des Usability Engineerings zählt die Beschreibung der Anwender und ihrer Umgebung, die Darstellung der Benutzerschnittstelle, die Methoden, um Risiken zu analysieren und zu minimieren sowie die Planung und Durchführung der eingesetzten Usability Methoden mitsamt der Ergebnisse [9].

In Europa gilt seit Mai 2017 die Medical Device Regulation (MDR) [10]. Dieses muss in nationales Recht umgewandelt werden, was in Deutschland durch das Medizinproduktegesetz (MPG) [1] geschieht. Am 26.05.2020 soll das MPG durch das Medizinprodukte-Durchführungsgesetz (MPDG) [3] ersetzt werden. Für den Bereich des Usability Engineerings gibt es in dem neuen Gesetz keine nennenswerten Änderungen. Usability wird in Deutschland ebenfalls für die Verhinderung von Gefahrensituationen eingesetzt und wird als Teil des Risikomanagements angesehen [3]. Die dokumentationspflichtigen Usability-Artefakten werden in der IEC 26466-1 [8] aufgegriffen und wurden bereits in Abschnitt 2.3 vorgestellt. Die IEC 26466-1 kann auch in Amerika als Leitfaden verwendet werden.

4 Traceability des Usability Engineerings in der Medizintechnik

Nachdem in Kapitel 3 beschrieben wurde, welche Rolle Usability Engineering in der Medizintechnik spielt, geht es in diesem Kapitel um den Einsatz von Traceability. In Abschnitt 4.1 werden die Nutzergruppen beschrieben, die Traceability in der Medizintechnik einsetzen können. Anschließend wird der Nutzen von Traceability im Usability Engineering in Abschnitt 4.2 herausgestellt und schließlich werden in Abschnitt 4.3 Voraussetzungen beschrieben, die eine Traceability-Software für das Usability Engineering erfüllen soll.

4.1 Nutzergruppen

Im Bereich der Medizintechnik vereinfacht Traceability Medizinprodukthersteller das Anforderungsmanagement. Einige Hersteller beschäftigen Requirements Engineers, die die Erstellung und Verwaltung der Anforderungen verantworten. Sie ermitteln die Bedürfnisse der Anwender und weiterer Stakeholder und vermitteln diese den Entwicklern. Außerdem prüfen Requirements Engineers die Einhaltung der Anforderungen. Das ist durch Traceability möglich, weil mit Traceability jederzeit erkennbar wird, welche Anforderungen bereit umgesetzt und getestet worden sind. Risikomanager konzentrieren sich auf die Risiken, die das Medizinprodukt mit sich bringt. Sie überwachen das Risikomanagement und die Methoden, die darin eingesetzt werden. Mit Traceability können Risikomanager nachvollziehen, woher Risiken stammen und ob diese schon geprüft worden sind. Des Weiteren können Qualitätsmanager in der Medizintechnik an Traceability des Usability Engineerings interessiert sein. Sie stellen sicher, dass alle Prozesse, die im Zusammenhang mit dem Produkt stehen im Unternehmen reibungslos ablaufen können. Um das zu überprüfen ist Traceability ebenfalls ein geeignetes Mittel. Die Aufgaben des Requirements Engineers, Risiko- und Qualitätsmanagers können auch die Projekt- oder Produktmanager umsetzen. Diese müssen nach Fertigstellung des Produkts außerdem nachweisen können, dass alle gesetzlichen Vorgaben eingehalten worden sind. Dafür benötigen sie eine übersichtliche Dokumentation aus der hervor geht, dass alle ermittelten Risiken geprüft worden sind. Schließlich sind Usability Engineers ebenfalls an Traceability interessiert. Sie arbeiten mit vielen Dokumenten, die unter anderen den Testaufbau, Ergebnisse und Änderungsvorschläge beinhalten. Mit Traceability können sie schnell überprüfen, welche Risiken noch nicht getestet worden sind oder nach einer Änderung erneut getestet werden müssen. [11]

4.2 Nutzen von Traceability im Usability Engineering

In manchen Projekten bekommt ein Usability-Team in der Medizintechnik nur die Testfälle, welche sie evaluieren sollen. Traceability ermöglicht es, beidseitige Beziehungen zwischen Anforderungen und Artefakten herzustellen [4]. So kann das Usability-Team nachvollziehen, woher die Testfälle stammen und welche Anforderungen damit überprüft werden sollen. Anschließend können sie eine Verknüpfung zu ihren Ergebnissen herstellen. Somit können Projekte strukturiert bearbeitet werden, weil jederzeit nachvollzogen werden kann, welche Anforderungen bereits umgesetzt, geprüft und gegebenenfalls verbessert worden sind. Dieser Punkt ist ebenfalls hilfreich, um die in Abschnitt 3.2 genannten gesetzlichen Vorlagen zu erfüllen. Durch Traceability entsteht der Beweis, dass das Produkt auf Risiken bei der Anwendung geprüft und diese bestmöglich beseitigt worden sind. Des Weiteren kann durch Traceability den Stakeholdern eines Projekts schnell verdeutlicht werden, welche Designentscheidung aus welchem Grund getroffen wird. [12].

Traceability bietet nicht nur die Möglichkeit die Artefakte des Usability Engineerings in der Medizintechnik miteinander in Verbindung zu setzen. Die Verbindungen können über den gesamten Entwicklungsprozess entstehen und beschränken sich nicht nur auf die Daten des Usability Engineerings [4]. Das bedeutet, dass technische und Usability-Artefakte nicht voneinander getrennt verwaltet werden müssen und so ein besserer Gesamtüberblick des Projektes entsteht. Dadurch unterstützt Traceability bei der Projektplanung, da der aktuelle Fortschritt der Entwicklung leicht ermittelt werden kann [13].

4.3 Einsatz von Traceability im Usability Engineering

Grundsätzlich kann Traceability im Usability Engineering genauso wie im Software Engineering eingesetzt werden. Das kann

beispielsweise mit einer Traceability-Matrix umgesetzt werden. Die Matrix ist in der Regel eine Tabelle, welche die Systemanforderungen in der ersten Zeile und die Artefakte in der ersten Spalte der Tabelle auflistet. Die Traces werden in Form von Markierungen an den Schnittpunkten zwischen Anforderung und Artefakt dargestellt [4]. Das ist jedoch nur für kleine Projekte sinnvoll, da die Tabelle mit vielen Anforderungen und Artefakten unübersichtlich wird. Im Usability Engineering gibt es eine Besonderheit bei den Artefakten. Die vorgeschriebenen Pflichtartefakte aus Abbildung 3 bringen bei der Ausarbeitung weitere Artefakte mit sich. Im Vergleich zu Artefakten des Software Engineering entstehen beim Usability Engineering spezielle Artefakte. Dazu zählen Ton- und Videoaufnahmen, Zeichnungen, Fotos und Prototypen, wie zum Beispiel Klick-Dummys [14]. Um Traces erzeugen zu können, ist es sinnvoll die Artefakte digital vorliegen. Das bedeutet, dass einige Artefakte zunächst eingescannt oder fotografiert werden müssen. Anschließend bietet es sich sowohl im Usability als auch im Software Engineering an, eine Traceability Software zu verwenden. Diese kann Traces und eine Traceability Matrix automatisch erstellen.

Eine geeignete Traceability-Software muss es ermöglichen, verschiedene Link-Arten zu verwenden, um die Artefakte in sinnvolle Beziehungen zu setzen. Dies fördert die strukturierte Vorgehensweise im Projekt. Außerdem muss die Software Artefakte aus unterschiedlichen Softwaresystemen und in unterschiedlichen Formaten verwalten und öffnen können. So können Anforderungen aus einem Textverarbeitungsprogramm mit Testergebnissen in einer Tabellenkalkulation verbunden werden. Das hat zusätzlich den Vorteil, dass Teams, die unterschiedliche Programme verwenden, ihre Ergebnisse untereinander verlinken können. Jedoch dürfen nicht nur Standardformate, wie zum Beispiel, Textdateien, berücksichtigt werden, sondern auch die oben genannten Formate wie

Ton- und Videoaufzeichnungen. Das bedeutet, dass die Traceability Software entsprechende Artefakte in der passenden Usability-Software öffnen können muss. Außerdem soll die Traceability-Software einen Überblick über die Traces geben können, damit unbearbeitete Aufgaben und der Projektfortschritt gemessen werden können. Das bedeutet, dass die Software in der Lage sein muss, Beziehungen auszuwerten und daraus ein Gesamtbild zu erstellen.

5 Zusammenfassung und Ausblick

Traceability bietet entscheidende Vorteile, um die Qualität des Usability Engineerings in der Medizintechnik zu verbessern. Dennoch ist der Einsatz von Traceability noch nicht weit verbreitet. Eine mögliche Ursache ist, dass im Usability Engineering viele unterschiedliche Werkzeuge verwendet werden, um Usability-Methoden zu planen, durchzuführen und zu dokumentieren. Die unterschiedlichen Dateiformate machen es schwer die entstehenden Artefakte sinnvoll miteinander zu verknüpfen. Dies ist jedoch notwendig, um die Sicherheit der Medizinprodukte nachweisen zu können. Laut gesetzlichen Vorgaben muss der Prozess des Usability Engineerings nachvollziehbar dokumentiert werden. Um dies zu gewährleisten eignet sich der Einsatz einer Traceability-Software.

In einer kommenden Masterarbeit sollen weitere Anforderungen ermittelt werden, die eine Traceability-Software erfüllen muss, damit sie sinnvoll im Usability Engineering in der Medizintechnik eingesetzt werden kann. Dazu wird ein Konzept für eine Traceability-Software entwickelt und diese anhand eines Prototypens evaluiert.

6 Literaturverzeichnis

- [1] Deutscher Bundestag, Gesetz über Medizinprodukte (Medizinproduktegesetz): MPG, 1994.

- [2] J. Harer, C. Baumgartner, Anforderungen an Medizinprodukte, Carl Hanser Verlag GmbH & Co. KG, München, 2018.
- [3] Deutscher Bundestag, Medizinprodukterecht-Durchführungsgesetz: MPDG, 2019.
- [4] J. Cleland-Huang, O. Gotel, A. Zisman (Eds.), Software and Systems Traceability, 1st ed., Springer Verlag London Limited, s.l., 2012.
- [5] DIN EN ISO 9241-110:2008-09, Ergonomie der Mensch-System-Interaktion_ - Teil_110: Grundsätze der Dialoggestaltung (ISO_9241-110:2006); Deutsche Fassung EN_ISO_9241-110:2006, Beuth Verlag GmbH, Berlin.
- [6] DIN EN ISO 9241-210:2011-01, Ergonomie der Mensch-System-Interaktion_ - Teil_210: Prozess zur Gestaltung gebrauchstauglicher interaktiver Systeme (ISO_9241-210:2010); Deutsche Fassung EN_ISO_9241-210:2010, Beuth Verlag GmbH, Berlin.
- [7] S. Klein, M. Kraft, U.M. Hölscher, Entwicklung und Zulassung von Medizinprodukten, in: K.-H. Grote, B. Bender, D. Göhlich, H. Dubbel (Eds.), Dubbel - Taschenbuch für den Maschinenbau, Springer Vieweg, Berlin, 2018, pp. 699–700.
- [8] DIN EN 62366-1:2017-07; VDE, Medizinprodukte - Teil 1: Anwendung der Gebrauchstauglichkeit auf Medizinprodukte (IEC 62366-1:2015 + COR1:2016); Deutsche Fassung EN 62366-1:2015 + AC:2015, Beuth, Berlin.
- [9] Food and Drug Administration, Code of Federal Regulations Title 21: 21 CFR, 2019.
- [10] Europäische Parlament und Rat der Europäischen Union, Medical Device Regulation: MDR, 2017.
- [11] H. Brandt-Pook, R. Kollmeier, Softwareentwicklung kompakt und verständlich: Wie Softwaresysteme entstehen, 2nd ed., Springer Vieweg, Wiesbaden, 2015.
- [12] G. Regan, F. Mc Caffery, K. Mc Daid, D. Flood, Medical device standards' requirements for traceability during the software development lifecycle and implementation of a traceability assessment model, Computer Standards & Interfaces 36 (2013) 3–9. <https://doi.org/10.1016/j.csi.2013.07.012>
- [13] V. Casey, F. Mc Caffery, A lightweight traceability assessment method for medical device software, J. Softw.: Evol. and Proc. 25 (2013) 363–372. <https://doi.org/10.1002/smr.571>.
- [14] M. Richter, M. D. Flückiger, Usability Engineering kompakt, Benutzbare Produkte gezielt entwickeln, 3. Auflage, Springer, Berlin, 2013.

Combining word embeddings and convolutional neural networks to detect duplicated questions

Yoan Dimitrov

Reutlingen University

Yoan.Dimitrov@Student.Reutlingen-University.de

Abstract

Detecting semantic similarities between sentences is still a challenge today due to the ambiguity of natural languages. In this work, we propose a simple approach to identifying semantically similar questions by combining the strengths of word embeddings and Convolutional Neural Networks (CNNs). In addition, we demonstrate how the cosine similarity metric can be used to effectively compare feature vectors. Our network is trained on the Quora dataset, which contains over 400k question pairs. We experiment with different embedding approaches such as Word2Vec, Fasttext, and Doc2Vec and investigate the effects these approaches have on model performance. Our model achieves competitive results on the Quora dataset and complements the well-established evidence that CNNs can be utilized for paraphrase recognition tasks.

Keywords

Natural Language Processing, Word Embeddings, Sentence Classification, Convolutional Neural Networks, Deep Learning

CCS Concepts

- Computing methodologies ~ Machine learning ~ Artificial intelligence ~ Natural language processing

1 Introduction

Over the past decade, numerous online community forums such as Stack Overflow and Quora have emerged, enabling users to ask domain-specific questions and obtain appropriate answers. Since many users share the same interests and often ask the same questions, these online forums face the challenge of identifying semantically similar or equivalent questions in order to avoid redundancies and improve their service quality and user experience [3]. Successfully measuring the semantic similarity between pairs of questions is, however, a very difficult hurdle to overcome, since natural languages are ambiguous at all major levels of linguistics. Two questions could vary in length, punctuation, sentence structure, and choice of words and still have the same intention. For example, consider the following two questions: *How old are you?* *What is your age?*

Supervisor: Prof. Dr. Benjamin Himpel
Reutlingen University
Benjamin.Himpel@Reutlingen-
University.de

Informatics Inside
Wissenschaftliche Vertiefungskonferenz
20. Mai 2020, Reutlingen University
Copyright 2020 Yoan Dimitrov

Both questions have the same intention and can be answered with the same answer, but they both have a different vocabulary and sentence structure. These aspects, although seemingly trivial, are not always easy to detect. Recent years have seen great progress in detecting semantically similar questions with deep learning. In particular, Recurrent Neural Networks (RNNs), which are inherently ideal for sequential data such as text, have achieved state-of-the-art results [6, 7, 9, 19, 20, 23].

In this work, we tackle the task of identifying semantically similar questions from the officially released Quora dataset. We adopt the question similarity definition suggested by Bogdonova et al. [3]:

“Two questions are considered to be semantically equivalent if they can be answered by the same answer.”

Contrary to popular approaches, we propose a simple and yet effective method based on word embeddings and a Convolutional Neural Network (CNN) architecture introduced by Yoon Kim [21]. Our experimental results are comparable to state-of-the-art and also shed light on the importance of cleaning noisy text data before creating word embeddings, as it drastically affects the performance of the model.

The outline of the paper is as follows. Section 2 gives a brief overview of related work on the detection of semantically equivalent questions. All details of our proposed model as well as the necessary data preprocessing steps are presented in section 3. Our model is evaluated in section 4 and compared with other approaches. Lastly, concluding remarks and future research directions are discussed in section 5.

2 Related work

There have been numerous academic contributions dealing with paraphrase detection and the detection of semantically equivalent questions. The following section

gives an overview of recent research conducted with the Quora dataset.

Wang et al. [23] were the first to publish impressive results on the detection of duplicated questions with the Quora dataset. They proposed a bilateral multi-perspective matching (BiMPM) model, which compares sentence embeddings P and Q from multiple perspectives and in both directions ($P \rightarrow Q$ and $P \leftarrow Q$) [23]. The matching vectors are then aggregated with a Long Short-Term Memory Network (LSTM). The authors achieved an impressive 88.17% classification accuracy [23]. Tomar et al. [7] also took on a similar LSTM approach, but used character-level n -gram embeddings and pretrained their decomposable attention (DECATT) model with noisy data from other domains in order to improve overall performance [7]. They managed to obtain an accuracy of 88.40%.

Homma, et al. employed a Siamese Gated Recurrent Unit (GRU) and applied data augmentation to their model for better results [20]. They achieved 85% accuracy on the Quora dataset. Furthermore, they also experimented with the following distances metrics to calculate the similarity between sentence vectors: Euclidian distance, Cosine distance, and weighted Manhattan distance. According to their studies, the weighted Manhattan distance performed best [20].

A relatively unique approach for detecting duplicated questions from the Quora dataset was demonstrated in [9] by Deudon. The author uses a Variational Autoencoder to derive the intentions of the questions as normal distributions. He argues that decomposing the representation of a sentence in a mean vector and a diagonal covariance matrix accounts for uncertainty and ambiguity in natural languages and is therefore more appropriate than single vectors which capture relationships between multiple words and phrases. The Gaussian intents are fed into a Variational Siamese network and compared using a Wasserstein

distance metric. He recorded an accuracy of 88.86% on the Quora dataset [9].

Gong et al. [19] introduced a Densely Interactive Inference Network (DIIN) to achieve high-level text comprehension by hierarchically extracting semantic features from interaction space. They make use of a CNN to acquire important features, which are then aggregated and passed to an output layer. The output layer predicts a confidence level for each class. Their model achieved state-of-the-art performance on multiple, well known datasets. On the Quora dataset, they were able to achieve an 89% accuracy [19].

Shen et al. conducted a comparative study between Simple Word-Embedding based Models (SWEMs) and RNN/CNN based models [6]. Their aim was to understand when and why simple pooling strategies such as max-pooling and hierarchical pooling can be applied to word embeddings without any further processing. They found that such strategies are sufficient in some cases of natural language understanding. They experimented with 17 different datasets. Their best model on the Quora dataset achieved an accuracy of 83.03% [6].

3 Approach

In the following sections, we present the implementation details and background of our proposed model. We particularly focus on the data preprocessing pipeline and the relevant operations and layers of our network architecture.

3.1 Problem formulation

Let $P = (p_1, \dots, p_M)$ and $Q = (q_1, \dots, q_N)$ be two questions consisting of M and N words, respectively [23]. Furthermore, let $y \in \{0, 1\}$ be the binary label indicating whether questions P and Q are duplicated or not. Our model's goal is to predict the y label that best reflects the semantic relationship between P and Q . As Wang et al. point out [23], the problem can also be seen as the estimation of the conditional probability $P(y | P, Q)$.

3.2 Text Preprocessing

The Quora dataset consists of 404,351 question pairs with positive and negative labels indicating whether or not two questions are duplicated. Roughly 37% of the question pairs are duplicates. The question pairs go through several preprocessing steps before being converted into numerical representations. First, every question is stripped of all punctuation and other non-alphanumeric characters. Then each question is tokenized with the Natural Language Toolkit (NLTK) tokenizer [13]. We decide not to remove stop words as they are considered to be semantically important in many cases. Furthermore, we do not discard rare words that occur only once or twice since they make up 43.21% of the text corpus. To train our neural network, all input questions need to have a fixed length. Therefore, we clip all sentences longer than 40 words. We decided on the length 40 after finding out that less than 1% of the questions contain more than 40 tokens.

In an attempt to fix numerous spelling errors which could negatively affect the performance of the model, we employ a pre-trained Word2Vec model [14, 15] and SymSpellpy. The Word2Vec model was trained on a Wikipedia dataset and therefore has a huge vocabulary. It is assumed that words which cannot be found by the Word2Vec model either contain spelling errors or are unknown terms. Such words are passed to SymSpellpy, a tool that corrects spelling errors and segments words accordingly if necessary. Words, which SymSpellpy cannot correct, are not discarded. We assume that such words are unknown names of people or simply outside the vocabulary (OOV). A similar approach can be found in [20].

3.3 Word Embeddings

The preprocessed questions are transformed into numerical representations before being used as input to our neural network. We experiment with three different embedding approaches: Word2Vec [14, 15], Fasttext [11]

and Doc2Vec [16]. In addition, we introduce a fourth approach by combining Word2Vec and Term Frequency-Inverse Document Frequency (TF-IDF) [4, 8].

Word2Vec was introduced by Mikolov et al. in 2013 [14, 15]. The authors proposed two closely related neural network architectures for learning word vector representations: Continuous Bag-Of-Words (CBOW) and Skip-Gram [14, 15]. CBOW attempts to learn the vector representation of a target word by taking some surrounding context words into account. Contrary to CBOW, Skip-gram predicts some corresponding context words for a given target word.

Bojanowski et al. [11] introduced Fasttext, an extension to Word2Vec, after recognizing that rare words are often poorly estimated or even regarded as insignificant to some algorithms. Instead of only learning vector representations on whole words in a text corpus like Word2Vec does, Fasttext is capable of learning the n -grams within a word [2, 11]. This, in turn, allows Fasttext to be more sensitive to OOV and rare words. The model first splits words into n -grams and then feeds the n -grams to a CBOW or Skip-Gram model in order to learn the embeddings.

Doc2Vec, also considered to be an extension to Word2Vec, was introduced by Le and Mikolov in 2014 [16]. This model attempts to learn how to map both words and entire sentences to an embedding space while capturing important semantics [16].

In this work, Word2Vec, Fasttext, and Doc2Vec are trained using the Skip-Gram architecture. This is due to the fact that Skip-Gram works well on smaller data sets and is more susceptible to rare words [14, 15]. Each model is trained for 30 epochs with the same parameters and generates 300-dimensional word vectors.

For the fourth embedding approach, we first calculate the TF-IDF score of every word in the text corpus. Frequent words like “the, a,

and, or”, which are found across all sentences, are penalized; simultaneously rare words across the whole text corpus receive high scores [4, 8]. We iterate through the vocabulary V , which represents all unique words in the dataset, and multiply the vector representation w of each word with its corresponding tf_idf -score as shown in Equation 1. The result is an embedding matrix $E \in \mathbb{R}^{n \times d}$, the rows of which are the TF-IDF-weighted word vectors across 300 dimensions.

$$E = \sum_{i=0}^V w_i \times tf_idf_i \quad (1)$$

3.4 Network Architecture

The network used to classify question pairs as duplicates or non-duplicates is largely based on the CNN architecture for sentence classification introduced by Yoon Kim [21] in 2014.

We begin with an embedding layer, which maps all words in a pair of questions $\{P, Q\}$ to their corresponding d -dimensional vector representations. This results in the matrices $P \in \mathbb{R}^{n \times d}$ and $Q \in \mathbb{R}^{n \times d}$, where n is the maximum (null-padded) sentence length, and d is the dimensionality of the vector representations [1, 21, 22]. Note that the word vector representations could come from any of the four embedding approaches mentioned previously. The matrices P and Q are then fed to separate 1d convolutional layers to extract different levels of semantic features for each question. To avoid repetition we will only explain how features are extracted from the matrix P , since the exact same procedure is used for matrix Q .

Given the matrix P , whose rows consist of a sequence of word representations, we denote $P[i : j]$ as a sub-matrix of P from row i to row j [16]. In order to extract features, we apply a filter $w \in \mathbb{R}^{h \times k}$ to a window of h words in the matrix P , while the width k of the filter remains fixed [17, 21, 22].

A feature map $c_i \in \mathbb{R}^{n-h+1}$ is generated by repeatedly applying the filter w on the sub-matrices of P :

$$c_i = \alpha(w \cdot P[i : i + h - 1] + b) \quad (2)$$

Here, $b \in \mathbb{R}$ is a bias term and $\alpha()$ is a non-linear activation function [21, 22]. In this work, we use the element-wise Rectified Linear Unit (ReLU) activation function. Generally, we apply five filter windows $\{2, 3, 4, 6, 8\}$ 200 times to matrices P and Q . This means each convolution layer outputs a matrix $c \in \mathbb{R}^{(n-h+1) \times 200}$. After generating the feature maps for all corresponding filter sizes, we then apply a global max-pooling operation (Eq. 3) as well as a global min-pooling operation (Eq. 4) over the matrix c [21]. Furthermore, we split the matrix c into i equal chunks, apply the global max-pooling operation to each chunk, and then concatenate their max-values (Eq. 5).

$$\begin{aligned} c_{max} &= \max(c) \\ \text{with } c_{max} &\in \mathbb{R}^{1 \times 200} \end{aligned} \quad (3)$$

$$\begin{aligned} c_{min} &= \min(c) \\ \text{with } c_{min} &\in \mathbb{R}^{1 \times 200} \end{aligned} \quad (4)$$

$$c_{max_i} = \max(z_1) \oplus \dots \oplus \max(z_i) \quad (5)$$

where z_1, \dots, z_i represent four equal sub-matrices $\in \mathbb{R}^{(n-h+1) \times 50}$ of the matrix c . The symbol \oplus is used to denote the concatenation operator.

As Kim mentions in [21], max-pooling captures the most important feature for every feature map. In the same manner, the min-pooling operation captures the least important feature for every feature map. We include this feature as well as the c_{max_i} features, as they capture additional important information about each question.

In order to find out if matrices Q and P are semantically equivalent or not, we compare

their respective c_{max} , c_{min} and c_{max_i} features, which are extracted from each feature map. More formally, we calculate their cosine similarities:

$$\cos(\theta) = \frac{f_1 \cdot f_2}{\|f_1\| \times \|f_2\|} + \lambda \quad (6)$$

where f_1 and f_2 represent the corresponding c_{max} , c_{min} or c_{max_i} values for matrices Q and P and λ is an additional, trainable parameter with a range from -1 to 1. Recall that our model uses five filters with varying window sizes to obtain multiple features for each question. Therefore, we compare the features f_1 and f_2 obtained from each filter in both directions ($f_1 \rightarrow f_2$ and $f_2 \leftarrow f_1$) [23]. After matching all feature vectors for Q and P , we concatenate their cosine similarity values as follows:

$$x = x_1 \oplus x_2 \oplus \dots \oplus x_n, \quad (7)$$

The concatenated vector x is then passed through a softmax function, which produces the distribution over the classification classes $\{0, 1\}$, from which the strongest class is outputted.

3.5 Network inputs

Originally, we designed the CNN architecture to compare two questions in the form of $d \times n$ matrices. After some experimentation, however, we decided to also explicitly include the unique words $\{P_u, Q_u\}$ and keywords $\{P_k, Q_k\}$ of each question pair as additional, embedded inputs to the network, resulting in three inputs for each question. Thus, we apply convolutional filters to the pairs $\{P, Q\}$, $\{P_u, Q_u\}$, and $\{P_k, Q_k\}$ and extract their max , max_i , and min features. Thereafter, we compare the features of P , P_k , and P_u with the features of Q , Q_k , and Q_u using the cosine similarity function (Eq. 6). The resulting vectors are then concatenated and fed to the softmax function to generate the final prediction. This approach is depicted in Figure 1. It should be noted that activation and regularization layers are omitted to avoid further complexity.

In order to find the unique words for every question, we simply retrieve the words that do not intersect in a question pair. The keyword extraction from each question is based on the words' tf_idf -scores. The three words with the highest tf_idf -scores in a question are selected as keywords. Since frequent words are penalized by the TF-IDF-algorithm and therefore have a lower score, they are usually not included as key words.

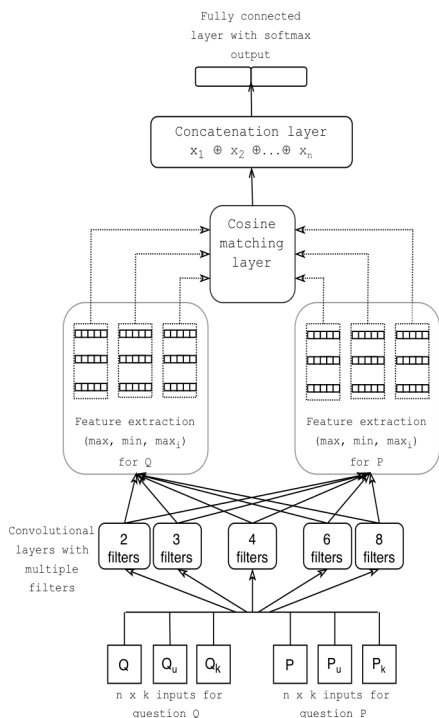


Figure 1: Proposed CNN architecture

3.6 Parameters and Training

For all experiments, we train with a mini-batch size of 64 and use the non-linear ReLU activation function on all convolutional layers. All the weights of these layers are randomly initialized with the *Xavier* initializer [18]. To avoid overfitting and stabilize training, we apply a dropout [10] rate of 0.1 as well as batch normalization with default parameters [12].

Our model is trained and evaluated on the exact same data partitions that Wang et al. published [23]. This allows us to compare our results fairly with theirs and other published results [6, 7, 9, 19, 20, 23], which also use the same data partitions. The dataset is split into three partitions: 10,000 question pairs for validation, 10,000 question pairs for testing, and the rest for training.

Our model's goal is to minimize the binary cross entropy of the predicted and true distributions. We employ the ADAM optimizer [5] with ($lr = 0.001$, $\beta_1 = 0.5$, $\beta_2 = 0.999$) to update model parameters. The learning rate lr is decayed after every two epochs by a factor of 0.1. During training, we also update the pre-trained word embeddings for the pairs $\{P, Q\}$ and $\{P_u, Q_u\}$. The weights of the keywords pair $\{P_k, Q_k\}$, are left static during training [21], since they do not improve model performance.

As we train our model with each of the four embedding approaches, we select the models that work best on the validation set and then evaluate them on the test set [23]. Accuracy is used as the evaluation metric for the dataset. The training is aborted if there is no update to the best accuracy on the validation set for the last 3 epochs.

4 Experimental Results

In Table 1 we compare our baselines with other models trained on the Quora dataset. Each model's validation accuracy (if published) and test accuracy are listed in the columns. The models are sorted by decreasing test accuracy. Note that all models with \diamond were reported by Wang et al. in their work [23]. Our baselines are denoted in bold in the last four rows of the table. Each baseline is labeled by the embedding approach used during training. Our experimental results surpass some models in terms of test accuracy and are also competitive with the state-of-the-art models. Among our models, we found Word2Vec to be the most consistent embedding approach.

Table 1: Performance comparison of different models on the Quora dataset

Model	Val Acc.	Test Acc.
DIIN [19]	89.44	89.06
VAR-Siamese [9]	89.05	88.86
DECATT [7]	88.89	88.40
BiMPM [23]	88.69	88.17
L.D.C. \diamond	-	85.55
Siamese-GRU [20]	-	85.0
Multi-Perspective-LSTM \diamond	-	83.21
SWEM-concat [6]	-	83.03
SWEM-aver [6]	-	82.68
Siamese-LSTM \diamond	-	82.58
SWEM-max [6]	-	82.20
Multi-Perspective-CNN \diamond	-	81.38
Siamese-CNN \diamond	-	79.60
W2V	87.45	86.73
D2V	87.29	86.52
FAST	87.35	86.59
W2V+TFIDF	87.03	86.02

In Table 2 we show several correct and incorrect label predictions on the validation set from our best model. The P column represents the predicted binary label, while the T column represents the true binary label from the validation set. The first two rows contain pairs of questions that our model classified correctly. The last three rows show questions that were incorrectly classified by our model.

During the training of our models we observed two interesting, behavioral patterns: A model sometimes tends to correctly predict most non-

uplicated pairs of questions, but at the same time has difficulties in predicting duplicates. The opposite case can also occur. In addition, we found that our models often predict question pairs correctly from a human point of view, but incorrectly according to the data set. Further investigation revealed that the Quora dataset has many question pairs that are labeled with a high degree of subjectivity. This evidently affects the performance of our model and its ability to generalize.

Table 2: Model prediction examples

Q1	Q2	P	T
How do I get funding for my web-based startup idea?	How do I get seed funding pre-product?	1	1
When did the first world war happen?	What were the causes of World War I?	0	0
How deep is the Titanic wreck?	How deep is Atlantic Ocean where Titanic sank?	0	1
What is the physical meaning of divergence, curl and gradient of a vector field?	Why curl represents rotation of a vector field?	0	1
What is definition of instrumentation?	What is biomedical instrumentation?	1	0

5 Conclusion

We presented a simple CNN approach built on top of word embeddings to detect duplicated questions from the Quora dataset. Correcting spelling errors in the data set with the pre-trained Word2Vec model improved our model performance by 3% on average. We employed four different embedding approaches, which

all led to similar results. Using global max-pooling, min-pooling and max-pooling on chunks to extract important features from each convolutional layer proved to be very effective. We also experimented with the average-pooling operation, but it performed poorly for our task. Overall, our competitive results show the potential of CNNs for the detection of semantically equivalent questions, a task that is typically solved by using RNNs.

With regard to future improvements of our approach, we believe using character-level embeddings [17] instead of word embeddings or even combining both methods could boost performance. Future work may also involve the study of other approaches for extracting features from convolutional layers, model ensemble techniques, and other similarity measures for comparing features.

6 References

- [1] A. Jacovi, O. S. Shalom, and Y. Goldberg. Understanding convolutional neural networks for text classification. In *Proc. EMNLP Blackbox NLP. ACL*. 2018.
- [2] A. Joulin, E. Grave, P. Bojanowski, and T. Mikolov. Bag of tricks for efficient text classification. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics (EACL)*. 2017.
- [3] D. Bogdanova, C. dos Santos, L. Barbosa, and B. Zadrozny. Detecting semantically equivalent questions in online user forums. *Proceedings of the 19th Conference on Computational Natural Language Learning*, 1:123–131, 2015.
- [4] D. Jurafsky and J. H. Martin. *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*. Pearson. Prentice Hall, Third Edition draft. 2018.
- [5] D. P. Kingma and J. Ba. Adam: A method for stochastic optimization. In *CoRR, abs/1412.6980*. 2014.
- [6] D. Shen, G. Wang, W. Wang, M. R. Min, Q. Su, Y. Zhang, C. Li, R. Henao, and L. Carin. Baseline needs more love: On simple word-embedding-based models and associated pooling mechanisms. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics*. 2018.
- [7] G. S. Tomar, T. Duque, O. Täckström, J. Uszkoreit, and D. Das. Neural paraphrase identification of questions with noisy pretraining. In *Proceedings of the 1st Workshop on Subword and Character Level Models in NLP*, page 142–147. 2017.
- [8] J. Ramos. Using tf-idf to determine word relevance in document queries. In *ICML*. 2003.
- [9] M. Deudon. Learning semantic similarity in a continuous space. In S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett, editors, *Advances in Neural Information Processing Systems 31*, pages 993–1004. Curran Associates, Inc. 2018.
- [10] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov. Dropout: A simple way to prevent neural networks from overfitting. *J. Mach. Learn. Res.*, 15(1):1929–1958. January 2014.
- [11] P. Bojanowski, E. Grave, A. Joulin, and T. Mikolov. Enriching word vectors with subword information. In *TACL 5*:135–146. 2017.
- [12] S. Ioffe and C. Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *ICML*. 2015.

- [13] S. Bird, E. Klein, and E. Loper. Natural language processing with Python: analyzing text with the natural language toolkit. *O'Reilly Media, Inc.*, 2009.
- [14] T. Mikolov, I. Sutskever, K. Chen, G. Corrado, and J. Dean. Distributed representations of words and phrases and their compositionality. *In Advances in Neural Information Processing Systems*. 2013a.
- [15] T. Mikolov, K. Chen, G. Corrado, and J. Dean. Efficient estimation of word representations in vector space. *In Proceedings of ICLR Workshop*. 2013.
- [16] V. Q. Le and T. Mikolov. Distributed Representations of Sentences and Documents. *In Proceedings of ICML*. 2014.
- [17] X. Zhang, J. Zhao, and Y. LeCun. Character-level convolutional networks for text classification. *In NIPS*. 2015.
- [18] X. Glorot and Y. Bengio. Understanding the difficulty of training deep feedforward neural networks. *In Proceedings of the thirteenth international conference on artificial intelligence and statistics, pages 249–256*. 2010.
- [19] Y. Gong, H. Luo, and J. Zhang. Natural language inference over interaction space. *In ICLR*. 2018.
- [20] Y. Homma, S. Sy, and C. Yeh. Detecting duplicate questions with deep learning. *In Proc. Int. Conf. Neural Inf. Process. Syst. (NIPS)*, pp. 1–8. 2017.
- [21] Y. Kim. Convolutional neural networks for sentence classification. *In Proceedings of the 2014 Conference on Empirical Methods for Natural Language Processing*. 2014.
- [22] Y. Zhang and B. Wallace. A sensitivity analysis of (and practitioners' guide to) convolutional neural networks for sentence classification. *arXiv preprint arXiv:1510.03820*. 2015.
- [23] Z. Wang, W. Hamza, and R. Florian. Bilateral multi-perspective matching for natural language sentences. *In Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence (IJCAI), pages 4144–4150*. 2017.

Visualisierung von neuronalen Netzen zur Unterstützung beim Training

Claudiu Bräuer
Hochschule Reutlingen

Claudiu.Braeuer@Student.Reutlingen-University.de

Abstract

In dieser Ausarbeitung wird auf Visualisierungsmöglichkeiten von neuronalen Netzen eingegangen. Ein neuronales Netz scheint zuerst nicht von außen einsehbar und ist somit für viele eine Blackbox. Häufig genutzte Python-Bibliotheken, zum Beispiel TensorFlow, werden vorgestellt und deren Stärken wie auch Schwächen präsentiert. Anhand dieser werden bereits bestehende Visualisierungen gezeigt und ihr derzeitiger Einsatz wird erläutert.

Durch einen Vergleich soll ersichtlich werden, welche Bibliothek am meisten Daten während des Trainings liefert, damit diese Informationen weiter verarbeitet werden. Diese Daten sollen so visualisiert werden, dass sie bei der Entwicklung eines neuronalen Netzes unterstützend sind. Ziel ist es, auf die Möglichkeiten einzugehen, welche geboten werden können. Durch eine Vereinfachung des Debuggings neuronaler Netze sollen weiterführende Entwicklungen in diese Richtung unterstützt werden.

Betreuer Hochschule: Prof. Dr. Uwe Kloos
Hochschule Reutlingen
Uwe.Kloos@Reutlingen-
University.de

Wissenschaftliche Vertiefungskonferenz
20. Mai 2020, Hochschule Reutlingen
Copyright 2020 Claudiu Bräuer

Schlüsselwörter

Python, Neuronale Netze, Machine Learning, Deep Learning, TensorFlow, The Microsoft Cognitive Toolkit, Visualisierung

CR-Kategorien

- Computing methodologies ~ Machine learning ~ Machine learning approaches ~ Neural networks
- Human-centered computing ~ Visualization

1 Einleitung

Seit Jahren sind künstliche neuronale Netze (KNN) in vielen Bereichen vertreten. Man kann sie beispielsweise in der Automobil-, Streaming- oder Medizinbranche finden. [1] Besonders ihre Flexibilität und ihr Umgang mit komplexen Problemen macht KNNs so wichtig. „Aufgrund der simplen Grundelemente der KNN können wir uns die Art und Weise, wie sie Daten verarbeiten, intuitiv vorstellen. Im Allgemeinen wird die Analogie zum Gehirn herangezogen.“ [2]

KNNs sind für Machine Learning und Deep Learning essenziell. Sie bestehen aus KNNs, wobei Netze für Deep Learning aus mehreren Hidden-Schichten bestehen, welche im Kapitel „Stand der Technik“ näher erläutert werden. „Das Ziel von Deep Learning ist es, Computer in die Lage zu versetzen bzw. sie lernen zu lassen, komplexe Aufgaben zu lösen, beispielsweise Tumore auf einem MRT-Bild zu erkennen, Menschen oder Gegenstän-

de in einem Video zu identifizieren, die Spracheingaben eines Smartphones zu interpretieren oder selbstständig ein Auto zu fahren.“ [1, S. 15] Deswegen wird in Kapitel 2 auf die Entwicklung von Deep Learning eingegangen. Dazu werden unterschiedliche Bibliotheken vorgestellt, welche die Entwicklung eines Deep-Learning-Netzes vereinfachen.

Im Kapitel „Analyse der Bibliotheken“ soll untersucht werden, wie viele Informationen während des Trainings eines KNNs entnommen werden können. Dabei soll vor allem auf die einzelnen Gewichte zwischen zwei Neuronen geachtet werden. Ziel ist es durch diese neuen Informationen, die der Entwickler erhält, das Debuggen zu unterstützen.

Nach der Untersuchung wird im Kapitel „Visualisierungen“ auf verschiedene Visualisierungsarten von Deep Learning Netzen eingegangen. Hierfür werden bereits bestehende und auch mögliche Neuarten vorgestellt. Durch eine Visualisierung können die neuronalen Netze für Entwickler verständlicher gemacht werden. Dadurch wird beispielsweise der Einblick in die Schichten und den Aufbau des neuronalen Netzes vereinfacht [1].

Zuletzt folgt eine kritische Betrachtung der Arbeit sowie ein Ausblick auf weitere Möglichkeiten.

2 Stand der Technik

2.1 Neuronale Netze

Neuronale Netze können für das maschinelle Lernen genutzt werden. Die Neuronen (auch Units genannt) dienen zur Eingabe, Verarbeitung und Ausgabe von Informationen [3]. Ein neuronales Netz wird in mehreren Schichten (auch Layer genannt) aufgebaut. Es gibt ein Input-Layer, das Informationen aufnimmt, Aktivitätslayer (auch Hidden-Layer), das der Informationsverarbeitung und der Gewich-

tung der Informationen dient, sowie ein Output-Layer, das die Informationen nach außen oder zur nächsten Input-Schicht trägt [3].

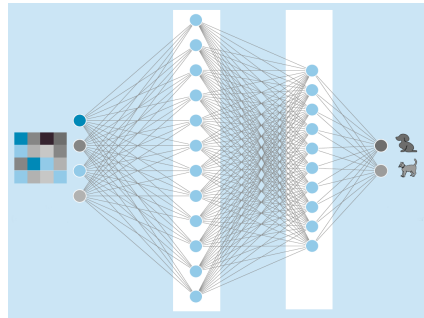


Abbildung 1: Struktur eines vollständig verbundenen (fully connected) Netzes mit zwei verdeckten Schichten [2, S. 62]

Die Units sind miteinander verbunden. Jede dieser Verbindungen besitzt ein Gewicht, wodurch die Units unterschiedlich stark beeinflusst werden. Handelt es sich um ein positives Gewicht, übt ein Neuron einen erregenden Einfluss auf ein anderes Neuron aus [3]. Wenn es sich um ein negatives Gewicht handelt, besitzt das Neuron einen hemmenden Einfluss und wenn die Verbindung ein Gewicht von Null besitzt, hat das Neuron keinen Einfluss auf das andere Neuron [3]. „Das Wissen eines neuronalen Netzes ist in seinen Gewichten gespeichert.“ [5] Das Lernen bei neuronalen Netzen wird „als Gewichtsveränderung zwischen den Einheiten definiert.“ [5]

Der Input eines Neurons wird aus dem Output des sendenden Neurons und dem Gewicht der Verbindung zusammengesetzt [6, S. 17]:

$$input_{ij} = a_j \cdot w_{ij}, \text{ wobei}$$

- i = empfangendes Neuron
- j = sendendes Neuron
- a_j = Output von j
- w_{ij} = Gewicht der Verbindung zwischen i und j , wobei das empfan-

gende Neuron i an erster Stelle geschrieben wird

Soll der komplette Input eines Neurons, auch Netzinput genannt, aus den einzelnen Inputwerten berechnet werden, wird eine Propagierungsfunktion benötigt [3]. Dafür wird häufig die Linearkombination verwendet. Dabei werden die einzelnen Inputwerte anhand der bereits genannten Formel berechnet, welche anschließend addiert werden [3]:

$netinput_i = \sum_j input_{ij} = \sum_j (a_j \cdot w_{ij})$, wobei die gleichen Bedingungen wie bei der Formel für die Berechnung des Inputs gelten.

Die Inputwerte eines Neurons können auch mehrdimensional sein und können somit einen Vektor X von Werten x_i sein, die einstellbare Gewichte besitzen. [2] „Die α Summe des Eingavektors mit den jeweiligen Gewichte bildet den Parameter für die anschließende Aktivierungsfunktion $\varphi(x)$. Der daraus berechnete Wert bildet [...] den Output] y des Neurons.“ [2, S. 59]

In Abbildung 2 ist eine Übersicht der Aktivierungsfunktionen dargestellt. Diese liefern je nach Eingabe unterschiedliche Ergebnisse. Ein Neuron wird als aktiviert angesehen, wenn die Eingabe den Schwellenwert überschreitet.

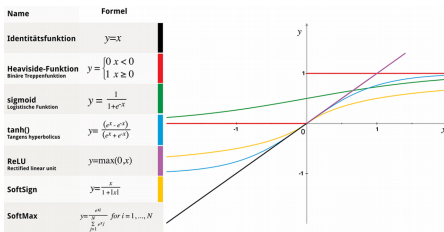


Abbildung 2: Übersicht von typischen Aktivierungsfunktionen [2, S. 60]

„Ein elementares Neuron berechnet eine lineare Funktion (z.B. $y = ax + b$). Das Lernen dieser linearen Funktion geschieht nun durch die Anpassung der Gewichte. Iterativ werden

die Eingabewerte eingespeist und die Berechnung der Ausgabewerte durchgeführt.“ [2, S. 60] Diese Berechnung wird wiederholt bis ein gewünschtes Ergebnis erreicht wird sofern möglich. Für die Anpassungen der Gewichte wird die Delta-Regel angewandt [2]:

$$w_{i, new} = w_{i, old} + \eta \cdot (y_i - \hat{y}_i) \cdot x_i$$

„Dabei sind y_i und \hat{y}_i die erwartete bzw. die berechnete Ausgabe des Perzeptrons und η ist die Lernrate.“ [2, S. 60] Als Perzeptron wird die Nervenzelle als Übertragung auf künstliche neuronale Netze bezeichnet [2]. Diese Berechnung wird in Abbildung 3 veranschaulicht.

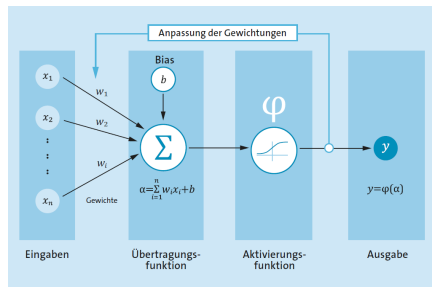


Abbildung 3: Schematische Darstellung des Prozesses zur Anpassung der Gewichte [2, S. 61]

Dieses Verfahren wird auch bei Deep Learning eingesetzt, wobei Deep Learning eine spezielle Optimierungsmethode von künstlichen neuronalen Netzwerken darstellt. Ein großer und entscheidender Unterschied zu herkömmlichen neuronalen Netzen ist die Komplexität der mittleren Schicht eines tiefen neuronalen Netzes. Sie besteht aus mehreren Hidden-Layers, welche für die Verarbeitung der Daten zuständig sind. „Herkömmliche neuronale Netze enthalten nur 2 bis 3 verborgene Schichten, während tiefe Netze bis zu 150 Schichten enthalten.“ [7]

Deep Learning ist in den letzten Jahren immer wichtiger geworden, da dadurch große Daten-

mengen verarbeitet und sehr genaue Ergebnisse geliefert werden können [7]. Außerdem ist es durch Deep Learning möglich, intellektuelle und mentale Prozesse darzustellen, von denen lange Zeit behauptet wurde, dass sie dem Menschen vorbehalten seien.

2.2 TensorFlow

„Tensoren sind die im Deep Learning übliche Erscheinungsform von Daten. Einfach ausgedrückt sind Tensoren nichts weiter als mehrdimensionale Felder, eine Erweiterung zweidimensionaler Tabellen (Matrizen) auf Daten mit mehr Dimensionen.“ [8, S. 5] Die Berechnungen werden in TensorFlow als Datenflussgraph abgebildet, sodass die Knoten des Graphs für die Operationen stehen und die Kanten für die Daten [8]. In Abbildung 4 wird dieses Verhalten bildlich veranschaulicht. „Die Daten fließen [...] durch einen Graphen von Rechenoperationen, aus denen unser Deep-Learning-Netz besteht.“ [8, S. 5]

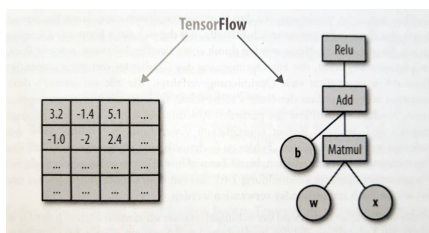


Abbildung 4: Ein Datenfluss-Berechnungsgraph [8, S. 5]

TensorFlow selbst arbeitet mit einer Kernkomponente, die auf C++ basiert. Diese besitzt zwei primäre Schnittstellen zur Hochsprache [8]. Die Python-Bibliothek ist die am weitesten entwickelte Schnittstelle dafür, die API kann direkt auf Hardwarekomponente zugreifen, was eine effiziente Ausführung auf einem laufenden Betriebssystem ermöglicht [8]. Dies kann mit der CPU, GPU oder mit beiden gleichzeitig Arbeiten, was bei der Installation entschieden werden muss. Mit Ke-

ras als festem Bestandteil von TensorFlow2 ist die Visualisierung von einzelnen inneren Eigenschaften besser möglich als mit TensorFlow, da hier die Visualisierung nur über TensorBoard möglich war. Mit TensorBoard alleine konnten beispielsweise nur Metriken des Netzes wie loss oder accuracy in Form von Kurvendiagrammen oder Histogrammen über die Zeit visualisiert werden. Durch die Integration von Keras können nun auch zusätzlich die Aktivierungen einzelner Layer als Bild visualisiert werden, siehe Abbildung 10 in Kapitel 4.2.

2.3 The Microsoft Cognitive Toolkit (CNTK)

Das Microsoft Cognitive Toolkit (CNTK) kann als Bibliothek für Python, C# und C++ verwendet werden oder in der Modellbeschreibungssprache BrainScript. Das Toolkit ist eines der ersten Deep-Learning-Toolkits, das das Open Neural-Network-Exchange-Format (ONNX) unterstützt. ONNX ist eine von Microsoft mitentwickelte Open-Source-Modelldarstellung zur Framework-Interoperabilität für gemeinsame Optimierungen [9]. Mit diesem Format können Entwickler Modelle zwischen Frameworks wie CNTK, Caffe2, MXNet und PyTorch verschieben [9]. ONNX dient auch der Visualisierung von Netzen und ihrem Inhalt. Dies kann bei der Entwicklung zur Unterstützung von Debugging genutzt werden.

CNTK-Eingänge, -Ausgänge und -Parameter sind als Tensoren organisiert. Jeder Tensor hat einen Rang: Ein Skalar ist ein Tensor von Rang 0, ein Vektor ist ein Tensor von Rang 1, eine Matrix ist ein Tensor von Rang 2 und so weiter. Diese unterschiedlichen Dimensionen werden als Achsen bezeichnet [10]. Jeder CNTK-Tensor hat einige statische und einige dynamische Achsen. Eine statische Achse hat während der gesamten Laufzeit die gleiche Länge. Bei den dynamischen können sich die

beinhaltenden Werte immer wieder ändern, wobei sich die Länge von Phase zu Phase ändern kann. Ihre Länge ist nicht im Voraus bekannt und kann so erst zur laufenden Phase erkannt werden, kann aber jederzeit abgerufen werden. Dieser Aufbau ähnelt dem TensorFlow-Verfahren, nur dass hier auch die Berechnungen in Tensoren abgelegt werden.

2.4 Visualisierung

Die Visualisierung wird zur Veranschaulichung von Netzdaten und Strukturen genutzt, sodass diese besser durch einen Entwickler interpretiert werden können. Außerdem ist es eine weitere Idee der Visualisierung dem Entwickler zusätzliche Randinformationen bereitzustellen.

In der Arbeit „GAN Lab: Understanding Complex Deep Generative Models using Interactive Visual Experimentation“ wurde ein interaktives Visualisierungstool vorgestellt, das „GAN Lab“ heißt. Mit diesem können Benutzer interaktiv generative Modelle trainieren und die Zwischenergebnisse des dynamischen Trainingsprozesses visualisieren. Dabei wird ein Generative Adversarial Network (GAN) genutzt. Das gesamte Tool wurde mit TensorFlow entwickelt und kann auf der Webseite der Entwickler genutzt und getestet werden. Ein GAN ist eine spezielle Art von Netz, das im Gegensatz zu anderen aus einer einfachen Eingabe eine komplizierte Ausgabe erzeugt. Dies ist in Abbildung 5 zu sehen. [11]

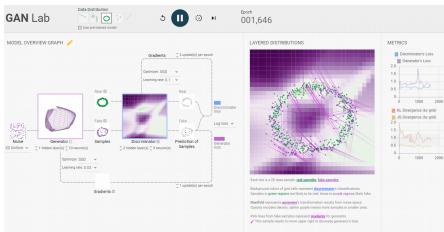


Abbildung 5: GAN Lab Webseite [11]

Hierbei wird zum Beispiel aus einem zweidimensionalen Bild eine dreidimensionale Figur erzeugt. Eine weitere Eigenschaft eines GAN ist die Rückkopplung fehlinterpretierter Daten an die Eingabeschicht, wodurch sich ein robusteres Lernverhalten ergibt. GAN Lab führt zudem neue interaktive Experimentierfunktionen zum Erlernen komplexer Deep-Learning-Modelle ein, z. B. schrittweises Training auf mehreren Abstraktionsebenen zum Verständnis komplexer Trainingsdynamiken. Dabei werden keine hoch komplexen Daten visualisiert, es wird mehr auf einer Verteilung von Punkten in nur zwei Dimensionen gearbeitet. Dadurch ist es aber immer noch möglich, die Mechanik des Systems zu zeigen. Zum einen sind Wahrscheinlichkeitsverteilungen im einfachen 2D-Raum (x, y) einfacher zu visualisieren als Verteilungen im Raum hochauflösender Bilder. Hier werden unterschiedliche Visualisierungsarten eingesetzt, um einen Überblick über das gesamte Verfahren zu erlangen. [11]

Ein weiterer Visualisierungsansatz wurde von Playground entwickelt. Dabei werden die Aktivierungen von Neuronen sowie deren Verbindungen beobachtet. In Abbildung 6 ist die interaktive Webseite von Playground zu sehen, auf der Visualisierungen wie bei GAN Lab live getestet werden können.

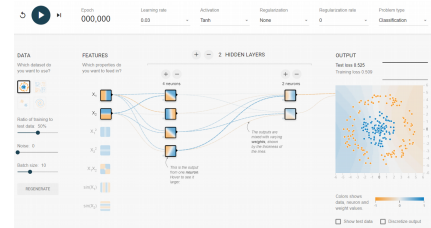


Abbildung 6: Ausschnitt der Playground-Webseite [12]

Dafür werden zweifarbige Datenpunkte in ein Netz eingespeist, wobei die orangefarbenen als negative Werte und die blauen als positive

Eingaben gewertet werden. Diese Einfärbung wird auch in den Verbindungen genutzt. In den verborgenen Schichten werden die Linien durch die Gewichte der Verbindungen zwischen Neuronen gefärbt. In der Ausgabeebene sind die Punkte je nach ihren ursprünglichen Werten orange oder blau gefärbt. Die Hintergrundfarbe zeigt an, welche Bereiche des Netzwerks als aktiv und welche als inaktiv vorhergesagt werden. Die Intensität der Farbe zeigt, wie sicher diese Vorhersage ist [12].

Die Visualisierung von Aktivierung in Hidden-Layers wurde auch in der Arbeit „Visualizing the Hidden Activity of Artificial Neural Networks“ untersucht. Hierbei wird nicht nur allein die Aktivierung von Neuronen visualisiert, sondern auch die Beziehungen zwischen dem Erlernten im Netz dargestellt. Dabei werden die Veränderungen sowohl innerhalb der Schichten als auch der Beobachtungen in Beziehung gesetzt. Zum einen zeigt die Visualisierung die Beziehung der einzelnen Epochen und die darin stattfindende Entwicklung. Zum Anderen werden Änderungen der Aktivierungen dieser Schichten infolge des Lernens gezeigt. Es wird also versucht der Zusammenhang zwischen dem Lernen und der Aktivierung zu visualisieren, um so ein besseres Verständnis hierfür zu gewinnen. In Abbildung 7 wird solch eine Visualisierung gezeigt [13].

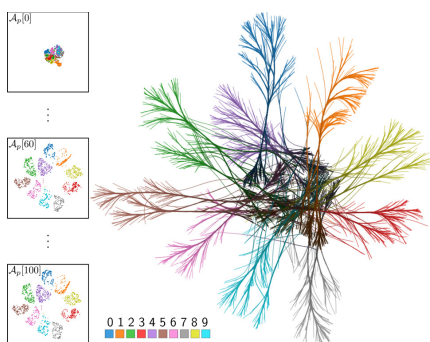


Abbildung 7: Visualisierung der Evolution zwischen den Epochen [13]

Eine weitere Visualisierung, die in der Arbeit dargestellt wurde, ist die Beziehung zwischen der Aktivierung einzelner Neuronen in Bezug auf die Klassifizierung von Daten. Dabei wurde eine Beziehung zwischen Klassifizierung von Daten und Aktivierung von Neuronen gefunden, durch die gezeigt werden konnte, welcher Bereich in einem Netz für die Erlernung einzelner Fragmente zuständig ist [13].

3 Analyse der Bibliotheken

In diesem Kapitel können nicht alle möglichen Entwicklungssysteme betrachtet werden. Außer den zwei Bibliotheken, die bereits im vorherigen Kapitel erwähnt wurden, gibt es noch viele andere Entwicklungssysteme wie zum Beispiel Caffe, Keras, Scikit-learn, Torch oder Theano. Keras ist ein Framework, das sich auf die Unterstützung verschiedener Backends stützt. Darunter sind auch CNTK und TensorFlow, weshalb Keras hier nicht separat betrachtet werden muss. Scikit-learn ist ein Machine-Learning-Framework für Python und keine Bibliothek [7]. Mithilfe der integrierten Bibliotheken kann ein Machine-Learning-Ansatz ohne hohen Aufwand realisiert werden. Torch ist auch ein Machine-Learning-Framework, das allerdings in Lua geschrieben wurde und mit CUDA von Nvidia arbeiten kann [7]. Durch die Erweiterung PyTorch kann es auch mit Python genutzt werden. PyTorch und Caffe wurden miteinander verbunden, weshalb beide gleich behandelt werden können. Theano ist im Gegensatz zu den eben genannten Systemen eine Python-Bibliothek, die allerdings nicht Open Source ist und deshalb nicht weiter berücksichtigt wird.

Bei der Untersuchung der Bibliotheken wurde berücksichtigt, wie deren Installation abläuft. Wichtig hierbei war, dass diese relativ einfach ist und auf vielen Plattformen unterstützt wird. So ist beispielsweise Caffe ungeeignet, da dessen Installation zu aufwendig ist. Dabei

muss die Kernkomponente von Caffe unabhängig der Bibliothek installiert werden. Wie bereits in Abschnitt 2 erwähnt, bietet TensorFlow zwei Installationsarten an. Für diese Arbeit wurde die GPU-Installation genutzt. Für die erfolgreiche Benutzung der GPU-Version musste ebenfalls CUDA, welches auch für die Nutzung von CNTK benötigt wird, installiert werden. Im Gegensatz zu TensorFlow arbeitet CNTK bereits standardmäßig mit der GPU, wobei das System selber entscheiden kann, ob die CPU oder GPU genutzt wird. Bei der Entwicklung von Testsystemen wurde festgestellt, dass die Fehlerbehandlung bei TensorFlow einfacher ist als bei CNTK. Zwar sind beide ähnlich ausführlich dokumentiert, allerdings ist die Fehlerausgabe bei TensorFlow deutlicher, da bei CNTK oftmals keine Fehlermeldungen erscheinen.

3.1 *Aufbau eines Netzes*

Das Erzeugen eines Netzes in TensorFlow und CNTK erfolgt über Python-Code. Dabei werden entweder Python-Dictionaries genutzt, die auch eine JSON Struktur aufnehmen und so ein Netz erstellen, oder Methoden, die direkt eine Aufbaustruktur bilden. In CNTK werden alle Methoden über die Klasse Layers aufgerufen. Jede Methode liefert einen vollständigen Zugriff auf das so erzeugte Layer. Bei TensorFlow können auch die Methoden der Klasse Layers genutzt werden, allerdings ist so kein Direktzugriff auf die beinhalteten Daten möglich. Über Placeholder kann das nachträglich erreicht werden oder wie im Codeblock 1 strukturiert werden.

```
#Aufbaustruktur in CNTK
conv1 = C.layers.Convolution2D((5,5),32,
                                pad=True)(input)
pool1 = C.layers.MaxPooling((3,3), (2,2))
      (conv1)
var.append(conv1.W)

#Aufbaustruktur in TensorFlow
layer_nodes = int(c_layer.get_shape()[1])
hidden_weights=tf.Variable(tf.truncated_normal(last_layer_nodes, hidden_nodes),
                           stddev=1./np.sqrt(last_layer_nodes)),
                           name='weights')

hidden_bias=tf.Variable(tf.constant(0.01,
                                     shape=(hidden_nodes,)), name='biases')

variables.append(hidden_weights)
```

Codeblock 1: Aufbau eines Netzes in CNTK und TensorFlow

In diesem Beispiel werden auch die Daten für mögliche Visualisierungen vorbereitet, sodass die Zugriffsvariablen in einem Array gespeichert werden, damit diese während des Trainings ausgelesen werden können. Bei der Betrachtung der beiden Bibliotheken wurde festgestellt, dass der Zugriff auf Information der Hidden-Layer bei CNTK beim Methodenaufruf zurückgegeben wird, wohingegen TensorFlow eine spezifische Abfrage benötigt. Hier hat sich der spezifische Zugriff von Vorteil erwiesen, da sonst unnötige Information weiter getragen werden. Ein weiteres Problem bei CNTK ist es, dass nicht direkt ersichtlich ist, um welche Art von Tensor es sich handelt. Dies liegt daran, dass in jedem weiteren zusätzlichen Layer die Variable mit aufgestapelt werden. Dies erweist sich beim weiteren Verlauf der visuellen Ausgabe als Nachteil.

3.2 *Das Trainieren von Netzen*

Für ein Training müssen erst Lernrate, Kostenfunktion und Metriken definiert werden. Die Kostenfunktion wird auch als Loss-Funktion bezeichnet, durch welche bestimmt wird, ob ein Trainingsdurchlauf erfolgreich ist oder nicht. Diese Funktionen werden von den Bibliotheken geliefert, können aber auch durch

eigene ersetzt werden. Für die Visualisierung von den Bibliotheken kann jeweils TensorBoard genutzt werden. Hierfür wird bei CNTK die Methode `TensorBoardProgressWriter()` verwendet und bei TensorFlow die Methode `tf.summary.FileWriter()`. Beide Methoden liefern einen `FileWriter`, der in einem externen Logfile die Informationen abspeichert, die von TensorBoard ausgelesen werden können und so Visualisierungen ermöglichen. Bei TensorFlow wird eine Session für das Training initialisiert und durch die Methode `run()` wird ein Durchlauf gestartet und währenddessen werden alle Berechnungen durchgeführt. Mit `w = session.run(variables[1])` werden die Gewichte vom ersten Hidden-Layer ausgelesen, wobei alle Gewichte vom Layer in einem mehrdimensionalen Array zurückgegeben werden. Bei CNTK wird über die Methode `Trainer()` das Training vorbereitet. Dabei wird der Methode das Modell, die Funktionen und der Lernparameter übergeben. Durch den Aufruf der `train_minibatch()`-Methode in der Epochenschleife wird immer ein Update durchgeführt. Dies verändert die Gewichte der Layer und kann dann über die übermittelte Variable im Array ausgelesen werden.

```
# Speicherung bei CNTK
for parameter in cnn.parameters:

    tensorboard_writer.write_value(parameter.
uid + "/mean",
reduce_mean(parameter).eval(), epoch)

# Speicherung bei TensorFlow
summaries = sess.run([accuracy_summary,
loss_summary],
feed_dict={images:img_input,
labels:labels_output})
for summary in summaries:
    summary_writer.add_summary(summary,
global_step=step)
```

Codeblock 2: Speicherung der Daten für Tensorboard

4 Visualisierungen

Ob ein bereits fertig trainiertes neuronales Netz noch verbessert werden kann, wird durch eine Visualisierung erkennbar gemacht. Dies ist wichtig, da bereits fertig trainierte Netze funktionsfähig sind und den Anschein erwecken, als wäre keine Verbesserung nötig. Visualisierungen ermöglichen ein Verständnis für das aufgebaute Netz, ohne dabei den Quellcode verstehen zu müssen.

4.1 Visualisierung eines Netzes

Eine Visualisierungsmöglichkeit ist die Visualisierung des Netzes. Ein Netz besitzt, wie in Kapitel 2.1 beschrieben, eine Graphenstruktur, weshalb das Netz durch eine Graphenvisualisierung veranschaulicht werden kann. In den Abbildungen 8 und 9 ist eine interaktive Netzvisualisierung zu sehen. Hierfür wurde der Code genutzt, der bereits in Kapitel 3.1 gezeigt wurde.

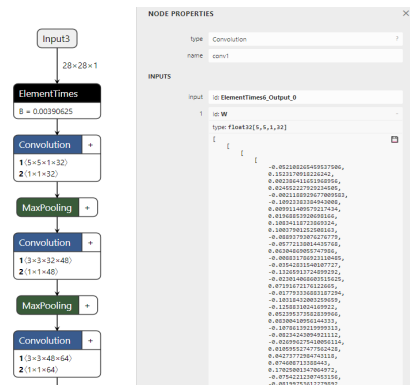


Abbildung 8: ONNX Darstellung von CNTK generiert

In Abbildung 8 ist ein Graph des Netzes mit Open Neural Network Exchange (ONNX) dargestellt. ONNX bietet beim Anwählen eines Knotens die Möglichkeit nähere Details von diesem zu erhalten. Dabei werden die konfigurierten Eigenschaften einer Methode

angezeigt. In der Abbildung befinden sich sowohl das Netz, als auch Detailinformation dessen, das mit CNTK erstellt wurde. Diese beiden arbeiten mit der ONNX-Konfiguration und können so darüber dargestellt werden. CNTK bietet die Möglichkeit ein Netz auch während des Trainings zu speichern, sodass bei jeder Erzeugung die Gewichte ausgegeben werden können. Dies ist in Abbildung 8 auf der rechten Seite zu sehen.

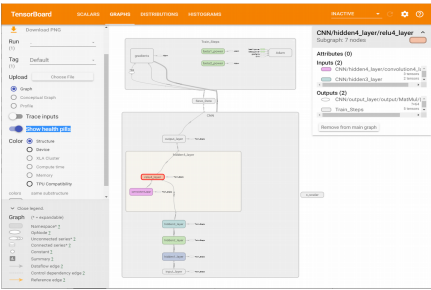


Abbildung 9: TensorBoard-Darstellung

In Abbildung 9 ist ein visualisiertes Netz mit TensorBoard zu sehen, welches von TensorFlow generiert wird. Mit diesem kann man beim Anklicken eines Knotens ebenfalls weitere Details einsehen und bei einem Doppelklick bietet es die Möglichkeit, in den Knoten tiefer einzusehen, sodass weitere Tensoren gezeigt werden. Durch die Eigenschaft des Eintauchens in die tiefere Struktur ist es möglich die Konfigurationen und Verknüpfungen der Tensoren, wie sie in Kapitel 2.3 vorgestellt wurden, zu sehen.

4.2 Visualisierung beim Training

Eine der bekanntesten Visualisierungen beim Training ist ein Kurvendiagramm der Loss- oder Wurfunktion. Dabei werden die einzelnen Schritte jeweils berechnet und beim Epochendurchlauf ausgegeben. In Abbildung 10 wird ein solches Diagramm gezeigt. Mithilfe

dieses Diagramms kann der Erfolgsstatus eines Trainings betrachtet werden.

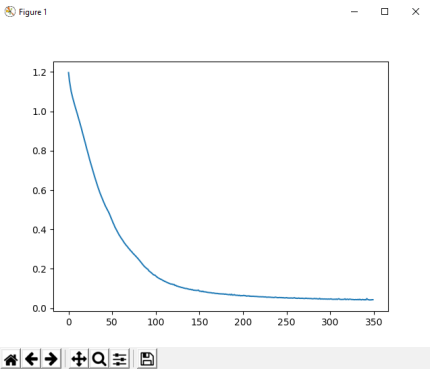


Abbildung 10: Diagrammdarstellung durch Python Graph Plotting

Diese Art der Visualisierung ist nicht spezifisch für eine Deep-Learning-Bibliothek, sondern kann auch durch die Python-eigene Plotting-Bibliothek ermöglicht werden.

Des Weiteren können auch die Aktivierungen eines jeden Layers ausgegeben werden. So eine Ausgabe ist auch mit TensorBoard möglich, was in Abbildung 11 gezeigt wird. Auch mit anderen Bibliotheken wie z. B. Keract oder PiPy kann eine solche Ausgabe erreicht werden. Bei dieser Art der Visualisierung werden die berechneten Ergebnisse in den Hidden-Layers als Bildmuster ausgegeben. Im Gegensatz zu den Visualisierungsansätzen in Abschnitt 2.4 wird hier der Versuch unternommen, ein visuelles Bild zu generieren, bei dem nur eine Aussage über ein gesamtes Layer getroffen wird und nicht auf die einzelnen Neuronen eingegangen wird. Bei den Aktivierungen von Layern werden die Funktionen, die in Abbildung 2 in Kapitel 2.1 gezeigt wurden, genutzt. Diese Funktionen dienen zur Bestimmung der Aktivierung des Neurons, wodurch Aussagen über die Weiterleitung von Daten zum nächsten Layer getroffen werden können.

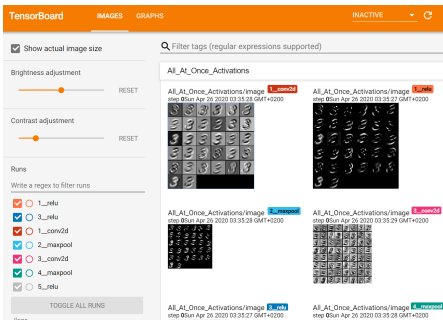


Abbildung 11: Visualisierung der Aktivierungen

In Abbildung 11 ist auf der linken Seite unter dem Punkt „Runs“ die jeweiligen Layerbezeichnungen zu sehen. Auf der rechten Seite ist die zugehörige Bezeichnung farblich zu sehen, wobei der aktuelle Lernschritt, sowie die jeweiligen Bilder der einzelnen Knoten eines Layers zu sehen sind.

4.3 Visualisierung der Gewichte

Bisher wurde auf viele unterschiedliche Arten der Visualisierung eingegangen, wobei unterschiedliche Aspekte eines Netzes visualisiert worden sind. Die meisten Ansätze sind bereits untersucht und werden im Herstellungsprozess eines Netzes als Unterstützung eingesetzt. Für eine Verbesserung dieser Prozesse gibt es einige neue Ansätze, wobei detailliertere Information vom Inneren eines Netzes erreicht werden sollen. Bei Playground werden erste neue Ansätze der Aktivierungen von Neuronen visualisiert. Die Ansätze in der Arbeit „Visualizing the Hidden Activity of Artificial Neural Networks“ wurden erweitert, indem die Epochen betrachtet wurden.

Bei beiden werden die Aktivierungen selbst näher untersucht, wobei die Aktivierungen nicht nur durch die Aktivierungsfunktionen erreicht werden, sondern auch durch die einzelnen Gewichte eines Neurons, was in Abbil-

dung 3 in Kapitel 2.1 gezeigt wurde. Auf der rechten Seite von Abbildung 8 wird ein kleiner Ausschnitt der Gewichte eines Convolution-Layers mit einer Größe von $5 \times 5 \times 1 \times 32$ dargestellt. Dies sind 800 Gewichte allein bei einer Trainingseinheit für ein einziges Layer und ein Deep-Learning-Netz besteht aus mehreren Hidden-Layers. Durch mehrere Layers, mehrere Epochen und große eingehende Lerndaten ist es nicht möglich die damit verbundenen Daten ohne großen Informationsverlust zu betrachten. Aufgrund dessen ist eine Visualisierung auch in Hinsicht auf die Informationsbeschaffung für Big Data sinnvoll. Zu viele Ausgaben oder zu lange Berechnungen während eines Trainings erhöhen die Trainingslaufzeit. Für erste Untersuchungen einer Visualisierung von Gewichten wurde zunächst Python Matplotlib genutzt. Das Netz, das untersucht wird, kann in Abbildung 8 gesehen werden. Dieses Netz wird mit dem MNIST-Datensatz trainiert, um eine Merkmalsklassifizierung zu erreichen.

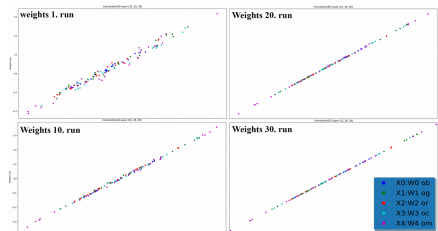


Abbildung 12: Visualisierungsdurchläufe für Gewichte

In Abbildung 12 können die Ergebnisse eines Durchlaufs betrachtet werden. Hierbei ist zu sehen, dass die Gewichte zunächst teilweise schwanken. Bereits nach zehn Durchläufen kann erkannt werden, dass sich alle Gewichte linear sortieren. Dies kann dafür sprechen, dass andere Visualisierungen möglicherweise ein detaillierteres Bild mit sich bringen. Nach der Untersuchung der Gewichte eines Neurons wird diese Annahme mehr bestätigt. In Abbildung 13 ist dieser Verlauf zu sehen. Da-

bei ist ersichtlich, dass sich diese nach einigen Durchläufen linear anordnen.

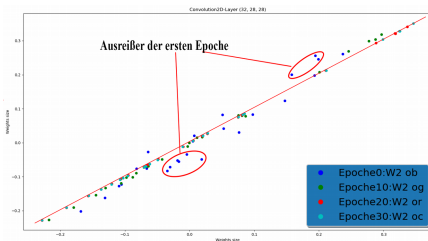


Abbildung 13: Gewichte eines Neurons über 30 Epochen

In einer weiteren vorgenommenen Betrachtung ist zu sehen, wie sich die Gewichte nach einem längeren Durchlauf verhalten. Hierbei wurden mehrere Iterationen in unterschiedlichen Konfigurationen ausgeführt. Dabei wurden sowohl CNTK als auch TensorFlow verwendet. Es sind keine ersichtlichen Unterschiede entstanden. In Abbildung 14 sind alle Gewichte des Layers aus dem gesamten Training abgebildet. In dieser Darstellung ist zu sehen, dass sich die meisten Daten zentral sammeln.

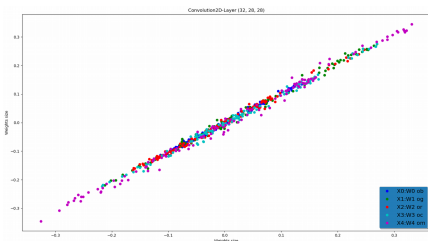


Abbildung 14: Alle Gewichte vom gesamten Training

Aus diesen Informationen können keine großen Rückschlüsse gezogen werden. Aus diesem Grund sollen die Gewichte auf andere Weisen visualisiert werden. Eine Möglichkeit wäre den in Kapitel 2.4 beschriebenen Clusteransatz zu verwenden. Die Visualisierung könnte dann ähnlich wie in Abbildung 7 aus-

sehen, wodurch mehr Informationen abgeleitet werden könnten.

5 Fazit und Ausblick

Für die Erstellung neuronaler Netze stehen mittlerweile viele unterschiedliche Frameworks und Bibliotheken zur Verfügung. Diese unterscheiden sich in der Art und Weise wie Netze aufgebaut werden sowie in den Tools für die Generierung. Bei der Untersuchung dieser ist aufgefallen, dass eine gut dokumentierte Bibliothek sehr wichtig ist. Bei TensorFlow und CNTK ist dies der Fall. Bei TensorFlow gibt es im Vergleich zu CNTK mehr Möglichkeiten, da es länger existiert und die entstandene Community sehr groß ist.

Da es für die Entwicklung von neuen Visualisierungen wichtig ist, dass die Daten, mit denen man arbeitet, sinnvoll sind, ist es wichtig ein funktionierendes Netz zu nutzen. Um ein funktionierendes Netz zu erstellen, können die bereits bestehenden Visualisierungen genutzt werden, da so gedebuggt werden kann. TensorFlow unterstützt die meisten Visualisierungsmöglichkeiten und eignet sich daher hierfür.

Anhand dieser Informationen ist der nächste Schritt mithilfe der gespeicherten Gewichte eine Visualisierungsart dafür auszuarbeiten. Dafür können unterschiedliche Visualisierungen eingesetzt werden, wie z. B. durch Punktwolken oder Streudiagramme. Auch der Ansatz, den Playground verfolgt hat, wäre denkbar. Die Idee wäre ein Tool zu entwickeln, bei dem es möglich ist, die Veränderungen von Gewichten bei einzelnen Layers einzusehen. Dadurch könnten mögliche Zusammenhänge vielleicht erkannt werden, die beim Entwickeln von neuen Netzen nützlich sind.

Da das Trainieren eines Netzes lange dauern kann, wäre eine weitere Idee dieses Tool nicht in Echtzeit zu verwenden, sondern eher das Auslesen der Gewichte in Etappen durchzu-

führen. So würden weniger Berechnungen dauerhaft stattfinden. Außerdem könnte ein Zeitstrahl visualisiert werden, in dem die Gewichte zu unterschiedlichen Zeiten sichtbar sind.

6 Literaturverzeichnis

- [1] M. Deru und A. Ndiaye, *Deep Learning mit TensorFlow, Keras und TensorFlow.js*. Rheinwerk Computing. 1. Auflage. Erstellt, 2019.
- [2] R. Schwaiger und J. Steinwendner, *Neuronale Netze programmieren mit Python*. Rheinwerk Computing. 1. Auflage. Erstellt, 2019.
- [3] U. Lämmel und J. Cleve, *Lehr- und Übungsbuch künstliche Intelligenz*. Fachbuchverl. Leipzig im Carl-Hanser-Verl., München. 2. Auflage. Erstellt, 2004.
- [4] R. D. Günter und F. Beck, *Neuronale Netze - Eine Einführung - Units*. Website 2018. Online verfügbar unter <http://www.neuralesnetz.de/units.html>. Besucht am 27.03.2020.
- [5] R. D. Günter und F. Beck, *Neuronale Netze - Eine Einführung - Verbindungen*. Website 2018. Online verfügbar unter <http://www.neuralesnetz.de/verbindungen.html>. Besucht am 27.03.2020.
- [6] G. D. Rey und K. F. Wender, *Neuronale Netze. Eine Einführung in die Grundlagen, Anwendungen und Datenauswertung*. Psychologie Lehrbuch. Huber, Bern. 1. Auflage. Erstellt, 2011.
- [7] The MathWorks, *Deep Learning: Drei Dinge, die Sie wissen sollten*. Website 2020. Online verfügbar unter <https://de.mathworks.com/discovery/deep-learning.html>. Besucht am 27.03.2020.
- [8] T. Hope, Y. S. Resheff, und I. Lieder, *Einführung in TensorFlow. Deep-Learning-Systeme programmieren, trainieren, skalieren und deployen*. Safari Tech Books Online. O'Reilly, Heidelberg. 1. Auflage. Erstellt, 2018.
- [9] Chrisbasoglu, *The Microsoft Cognitive Toolkit - Cognitive Toolkit - CNTK*. Website 2020. Online verfügbar unter <https://docs.microsoft.com/de-de/cognitive-toolkit/>. Besucht am 27.03.2020.
- [10] Microsoft, *Working with Sequences — Python API for CNTK 2.6 documentation*. Website 2018. Online verfügbar unter <https://cntk.ai/pythondocs/sequence.html>. Besucht am 27.03.2020.
- [11] Kahng, M., Thorat, N., Chau, D. H. P., Viegas, F. B., and Wattenberg, M. Website 2018. GAN Lab: Understanding Complex Deep Generative Models using Interactive Visual Experimentation. IEEE transactions on visualization and computer graphics.
- [12] D. Smilkov und C. Shan, *Tensorflow — Neural Network Playground*. Website 2017. Online verfügbar unter <http://playground.tensorflow.org/>. Besucht am 19.03.2020.
- [13] Rauber, P. E., Fadel, S. G., Falcao, A. X., and Telea, A. C. Website 2017. Visualizing the Hidden Activity of Artificial Neural Networks. IEEE transactions on visualization and computer graphics 23, 1, 101–110.

Testumgebungen für Augmented Reality zur Anzeige autonomer Fahrfunktionen

Markus Rein

Hochschule Reutlingen

Markus.Rein@Student.Reutlingen-University.de

Abstract

In dieser Arbeit werden drei verschiedene Testumgebungen vorgestellt, welche in ein iteratives Vorgehen einfließen, um die Entwicklung von Augmented-Reality-Anwendungen zur Darstellung von autonomen Fahrfunktionen zu unterstützen.

Gestaltungsentwürfe und Softwareentwicklungen können in den Testumgebungen für unterschiedliche Zielsetzungen von Personenbefragungen vorgestellt und bewertet werden. Das entwicklungsbegleitende Testen ermöglicht eine frühzeitige Identifizierung von Änderungshinweisen, welche für einen gültigen Lösungsentwurf eingearbeitet werden können. Die entwickelten Testumgebungen sind ein verkleinertes Modell, ein Fahrsimulator und ein reales Fahrzeug. Eigenschaften, Funktionen und Aufbauten resultieren aus Erkenntnissen der Literatur und Erfahrungen aus ersten Entwicklungen. Diese und die Einsatzmöglichkeiten werden mit dieser Arbeit aufgezeigt.

Schlüsselwörter

Augmented Reality, Connected Automated Driving, Iterative Entwicklung, Testumgebung

Betreuer Hochschule: Prof. Dr.-Ing. Cristóbal Curio
Hochschule Reutlingen
Cristobal.Curio@Reutlingen-University.de

Betreuer Firma: Tobias Müller, M. Sc.
Robert Bosch GmbH
Tobias.Mueller8@de.bosch.com

Wissenschaftliche Vertiefungskonferenz
20. Mai 2020, Hochschule Reutlingen
Copyright 2020 Markus Rein

CR-Kategorien

• Human-centered computing • Software and its engineering ~ Software creation and management

1 Einleitung

Diese Arbeit stellt drei Testumgebungen vor, welche in ein iteratives Entwicklungsvorgehen einfließen können. Unterstützt werden kann damit der Entwurf und die Entwicklung von *Augmented Reality (AR)* Anwendungen zur Darstellung von Fahrfunktionen des sogenannten Connected Automated Driving.

Einem automatisierten Fahrzeug werden während der Fahrt Informationen von externen Informationsquellen über mobile Netzwerke übertragen, welche entsprechende Fahrentscheidungen beeinflussen. Dem Fahrgast können diese Informationen zum Nachvollziehen der Entscheidungen mit entsprechenden Visualisierungen auf AR-Anzeigegeräte veranschaulicht werden.

Zu diesen zählt beispielsweise das Anzeigen von zukünftigen Fahrmanövern, welche sich auf die physische Umgebung beziehen. Mit AR kann eine Visualisierung in die dreidimensionale Umgebung integriert werden, um einen örtlichen Bezug der Information herstellen zu können.

Geeignete Fahrinformationen zur Anzeige mit AR haben vorangehende Arbeiten bei der Robert Bosch GmbH untersucht. Der Autor dieser Arbeit entwirft und implementiert darauf aufbauend entsprechende Darstellungsformen bestehender Informationen. Die Entwicklung enthält Herausforderungen,

welche in vorangehenden Arbeiten bei Bosch aufgefallen sind.

Die Erarbeitung von AR-Darstellungslösungen sind von verschiedenen Einflüssen abhängig, sodass die Entwürfe und Implementierungen stetig während der Entwicklung mit dem Anzeigegerät kontrolliert werden müssen. Dies dient zur Prüfung, ob ein Anzeigeelement wie im Lösungsentwurf dargestellt wird.

Um Aussagen auf potenzielle Nutzer zu erfahren, sind ausgewählte Anzeigekonzepte prototypisch umgesetzt und involvierten Personen im realen Fahrzeug vorgestellt worden. Der Austausch hat sich während der ersten Entwicklungsphase als sehr wichtig erwiesen, da dadurch viele Verbesserungsvorschläge resultieren und Fehler detektiert werden.

Es hat sich jedoch herausgestellt, dass die Vorbereitung und Durchführung von Tests im realen Fahrzeug sehr aufwendig und situationsabhängig sind. Es müssen Personen vom Entwicklungsstandort zu einem Testabschnitt gebracht werden. Das hat einen Zeitaufwand zur Folge, da öfters zum Straßenabschnitt gefahren werden muss, um mit einer Vielzahl von Personen einen Test durchzuführen. Somit ist das Erfahren von Bewertungen und die Einarbeitung von Verbesserungsvorschlägen ein aufwendiger Prozess.

Am Teststraßenabschnitt angekommen setzten Personen im Fahrzeug die HoloLens auf und je nach Situation blendet die Anwendung Visualisierungen ein. Für entsprechende Anzeigen müssen wiederholt Situationen nachgestellt werden. Des Weiteren treten auch sicherheitskritische Situationen im Straßenverkehr auf, welche die Durchführung der Tests nicht gestatten.

Für die Entwicklung weiterer Visualisierungsentwürfe werden mit dieser Arbeit aufgrund der Einschränkungen der Testumgebung im realen Fahrzeug Alternativen erarbeitet. Es sollte die Möglichkeit bestehen, dass sich die im Projekt involvierten Personen stetig über erste Entwicklungserfolge austauschen und diese von potenziellen Nutzern bewertet

werden können. Einzelne Entwicklungsiterationen umfassen die Vorstellung und Bewertung von Visualisierungsentwürfen in den Testumgebungen und die Verfeinerung anhand der Ergebnisse.

1.1 Aufbau

Aufbauend zu den Kontextinformationen des Projektumfelds und -vorhabens der Arbeit, wird nachfolgend in Abschnitt 2 auf das Themengebiet AR eingegangen und ein Überblick des Anzeigegeräts *Microsoft HoloLens* gegeben. In Abschnitt 3 sind Testumgebungen mit AR aus wissenschaftlichen Arbeiten aufgeführt, welche Einfluss auf die eigene Entwicklung haben. In Abschnitt 4 werden iterative Entwicklungsvorgehen unterschiedlicher Fachrichtungen aufgezeigt, in welchen Testumgebungen zur Gestaltungs- und Entwicklungsbewertung eingesetzt werden können. Drei eigene Entwicklungen von Testumgebungen für den Einsatz im Projektkontext sind in Abschnitt 5 aufgeführt. Die Arbeit wird in Abschnitt 6 zusammengefasst und es werden zukünftige Weiterentwicklungsmöglichkeiten beschrieben.

2 Augmented Reality

Der Begriff Augmented Reality steht für ein Themengebiet, welches sich damit beschäftigt die Wahrnehmung eines Menschen mit virtuellen Informationen zu erweitern. Dieses hat sich in den letzten Jahrzehnten stetig durch Technologieentwicklungen und Erkenntnissen zur Anwendung weiterentwickelt.

Ein allgemeingültiges Verständnis des Themengebiets, unabhängig der verwendeten Technologie und Einsatz in einem Anwendungsgebiet, verdeutlicht das *Virtuality Continuum* nach Milgram et al. [6]. Erweitert wird dieses mit drei repräsentierenden Merkmalen eines AR-Systems, nach Azuma [1].

Das Kontinuum, welches in Abbildung 1 illustriert ist, beschreibt ein Mixed-Reality-Spektrum von der realen zur virtuellen Umgebung. Die reale Umgebung ist die, in welcher wir Menschen uns bewegen und wir mit unseren Sinnen wahrnehmen können. Dagegen ist die virtuelle Umgebung rein künstlich, die von

einem anderen Themengebiet der *Virtual Reality (VR)* bekannt ist. AR ordnet sich links in das Mixed-Reality-Spektrum ein und es ermöglicht, die Wahrnehmung der realen Umgebung mit virtuellen Informationen zu ergänzen, [1].

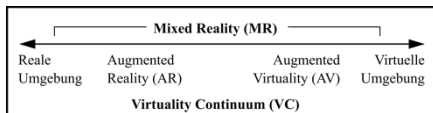


Abbildung 1: Virtualität-Kontinuum, Eigene Darstellung auf Basis von Milgram et al. [6]

Ein System nennt sich AR, wenn es die nachfolgende drei Merkmale nach Azuma [1] aufweist:

- Es kombiniert das Reale mit dem Virtuellen.
- Es ist interaktiv in Echtzeit.
- Es werden drei Dimensionen registriert.

Im Projektkontext dieser Arbeit wird die visuelle Wahrnehmung mit AR erweitert, indem einer Person mit einem entsprechenden Anzeigegerät Visualisierungen in die Sicht der physischen Umgebung eingeblendet werden. Es existieren verschiedenen AR-Anzeigearten, welche Van Krevelen und Poelman [11] beschreiben. Den größten Nutzen für diese Arbeit bieten sogenannte *see-through Head-Mounted Displays*. Diese sind Anzeigegeräte, welche auf dem Kopf getragen werden und visuelle Inhalte auf eine durchschaubare Projektionsfläche projizieren können. Diese Technologie ermöglicht einem Nutzer die Erweiterung des realen Sichtfelds und eine natürliche Interaktion im 3D-Raum.

Bereitgestellt wird diese Anzeigetechnologie unter anderem von einem Anzeigegerät namens *Microsoft HoloLens* der ersten Generation. Dieses ist ein mobiles Gerät, ohne eine Verbindung zu externen Power- und Recheneinheiten zu fordern. Es können 3D-Visualisierungen orts- und kontextgebunden in der dreidimensionalen physischen Umgebung angezeigt werden. In ihr kann man sich natürlich bewegen und mit virtuellen Objekten

interagieren. Diese können von verschiedenen Perspektiven durch Veränderung der eigenen Position und Kopfausrichtung betrachtet werden. Es ist ebenfalls möglich mit diesem Perspektivwechsel Informationen außerhalb des Sichtfelds zu erfahren.

3 Testumgebungen für Augmented Reality

Bei der Recherche von wissenschaftlichen Arbeiten sind keine Beschreibungen von Testumgebungen gefunden worden, welche auf den konkreten Schwerpunkt von AR zur Darstellung von automatisierten Fahrfunktionen mit AR-HMD eingehen. Der Autor dieser Arbeit ordnet vorgeschlagene AR-Testumgebungen in zwei Kategorien ein.

Ein Teil der vorgestellten Testumgebungen ermöglichen die Überprüfung, ob sich AR für einen bestimmten Anwendungsfall eignet (Kategorie 1). Beispielsweise stellen Nawab et al. [7] eine Testumgebung vor, um menschliche Faktoren beim Erledigen einer bestimmten Aufgabe mit AR zu untersuchen. Dem Nutzer werden AR-Visualisierungen angezeigt, um auf einen bestimmten Punkt im Raum hinzuweisen.

Andere Testumgebungen bieten die Möglichkeit AR Entwicklungen, wie eine bestimmte Funktion oder Visualisierung, testen zu können (Kategorie 2). In der Arbeit von Dias et al. [2] wird eine Testumgebung beschrieben, mit welcher ein neues Interaktionskonzept mit einem 3D-Modell auf die Benutzerfreundlichkeit getestet werden kann. Sauer et al. [10] beschreiben eine Testumgebung, mit welcher definierte Visualisierungen auf einen Tisch projiziert werden, um diese in AR zu erkunden und zugleich mit virtuellen und realen Objekten zu interagieren.

Wie in Abschnitt 1 erwähnt, können AR-Testumgebungen im Projektkontext eingesetzt werden, um Entwürfe und Entwicklungen von Informationsvisualisierungen zu bewerten. Diese Zielsetzung ordnet sich in die zweite Kategorie, sodass die vorgestellten Testumgebungen näher betrachtet werden.

Um ein neues Interaktionskonzept zu testen, benutzen Dias et al. [2] Marker, welche an den Fingern positioniert werden, um ein auf einen Marker projiziertes 3D-Objekt manipulieren zu können. Marker sind in dieser Arbeit dem System bekannte ausgedruckte Bilder. Ein Kamerabild wird auf die Erkennung eines Markers untersucht, welche entsprechend Programmfunktionen abrufen. Der Ansatz des Marker-Trackings zur Demonstration und für Tests von AR-Visualisierungen ist auch von Sauer et al. [10] durch eine Projektion auf einen Tisch verwendet worden. Für diese Arbeit eignet sich möglicherweise das Marker-Tracking, um die erstellten Visualisierungen auf einem Tisch anzuzeigen. Beispielsweise kann der Entwickler am Arbeitsplatz die erstellten Visualisierungen mit der HoloLens anschauen, ob diese wie gewünscht angezeigt werden.

Walch et al. [12] verwenden einen Aufbau einer Testumgebung in einem Fahrsimulator und die Darstellung von Informationsvisualisierungen in der VR, um Fahrentscheidungen in bestimmten Situationen von den Testprobanden zu erfahren. Der beschriebene Nutzen von VR kann jedoch nicht auf diese Arbeit übertragen werden, da die Visualisierungen sich in die reale und nicht in eine künstliche Umgebung einordnen. Jedoch kann von Walch et al. [12] das Testsetup übernommen werden, dass sich der Nutzer während einem Nutzertest in einer dem Auto ähnelnden Umgebung befindet. Somit wird eine realitätsnahe Umgebung für den Nutzer geschaffen und die Ergebnisse lassen sich möglicherweise auf den Einsatz im realen Fahrzeug übertragen.

Ergänzend verwendet Pfannmüller [9] eine Testumgebung im Fahrsimulator, um AR-Visualisierungskonzepte und -variationen von Nutzern bewerten zu können. Dabei wird ein Video einer Straßensituation auf einem Bildschirm eingeblendet und das Sichtfeld des Nutzers mit Visualisierungen von einem im Fahrzeug integrierten AR-Anzeigegerät erweitert. Für die Übertragbarkeit auf den Einsatz im physischen Straßenverkehr ist zusätzlich ein zweiter Test in der Umgebung des fahrenden realen Fahrzeugs durchgeführt worden.

Daraus lässt sich schließen, dass sich Voruntersuchungen in einer simulierten Umgebung eignen, um eine bestimmte Situation wiederholt nachzustellen. Jedoch beinhalten die Umgebung des realen Straßenverkehrs Einflüsse, welche nicht simuliert werden können.

4 Iterative Entwicklung

Für die Entwicklung von Anwendungen zur Anzeige autonomer Fahrfunktionen wählt der Autor dieser Arbeit ein iteratives Vorgehen. Dieses ist in den Grundsätzen der ISO Norm 9241–210 [3] enthalten, welche Empfehlungen zu menschenzentrierten Gestaltungsgrundsätzen und -aktivitäten zur Entwicklung interaktiver Systeme darlegt.

Ein iteratives Vorgehen bedeutet „[...] das Wiederholen einer Folge von Schritten, solange bis das gewünschte Ergebnis erzielt wurde“ [3]. Es besteht die Möglichkeit schrittweise Unklarheiten der Entwicklung zu beseitigen, indem Lösungsentwürfe in Form von Prototypen im Entwicklungsteam und mit potenziellen Benutzern ausdiskutiert werden. Verfeinerungen und Anpassungen von Gestaltungslösungen resultieren fortlaufend aus benutzerzentrierten Evaluationen und potenzielle Benutzer werden in die Gestaltung und Entwicklung miteinbezogen. [3]

In Abbildung 2 sind die einzelnen Aktivitäten der ISO 9241–210 in entsprechender Anordnung und Referenz dargestellt. Die Planung, von der das iterative Vorgehen ausgeht, ist eine essenzielle Aktivität, um für den Nutzer ein sinnvolles und nützliches interaktives System entwickeln zu können. Darauf folgend beginnt eine Iteration mit dem Verständnis und Beschreibung des Nutzerkontextes. Daraus lassen sich Anforderungen an die entwickelte Gestaltungslösung spezifizieren. Anschließend wird die Lösung evaluiert. Anschließend wird die Lösung evaluiert. Entsprechend der gewonnenen Ergebnisse wird entschieden, ob und von welcher ausgehenden Aktivität eine neue Iteration durchlaufen werden muss. Ein weiteres Ergebnis ist die vollständige Erfüllung der Anforderungen, was das Ende der Iterationen bedeutet, da eine gültige Lösung entwickelt worden ist. [3]

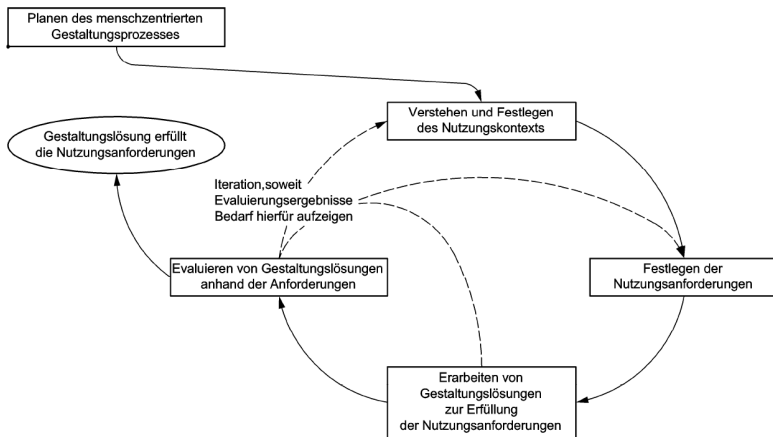


Abbildung 2: Wechselseitige Abhängigkeit menschzentrierter Gestaltungsaktivitäten [3, Bild 1]

Im Projektkontext dieser Arbeit stehen für einen Lösungsentwurf einer Visualisierung Gestaltungsaufgaben mit denen der Softwareentwicklung in Beziehung. Wie in den Grundsätzen der ISO 9241–210, werden in die Entwicklung fachübergreifende Kenntnisse und Perspektiven einbezogen, [3]. Es gelten Anforderungen und Ansichten zum Lösungsentwurf, die sich je nach Fachrichtung unterscheiden oder sich während der Entwicklung ändern können. Ein iteratives Vorgehen ermöglicht es, entwicklungsbegleitende Unklarheiten zu beseitigen. Zusätzlich können Anforderungen und Entwicklungen dynamisch angepasst werden. Intern in den jeweiligen Fachrichtungen hat sich ein iteratives Vorgehen etabliert, um gültige Entwicklungserfolge auf Anforderungserfüllung zu testen und Änderungen gemeinsam im Team auszuarbeiten.

Der Einsatz des iterativen Vorgehens in den einzelnen Fachrichtungen wird in den nachfolgenden Unterabschnitten aufgeführt. Zusätzlich wird ein Bezug zu dieser Arbeit hergestellt, wie eine Testumgebung als Präsentations-, Kommunikations- oder Bewertungsmedium eingesetzt werden kann.

4.1 Iterative Entwicklung im User Interface Design

Das iterative Vorgehen hat sich im Entwicklungsbereich von Designlösungen für Benutzeroberflächen etabliert. Dies wird unter anderem im oben erwähnten ISO Standard 9241–210 aufgeführt. Projektinvolvierten Personen werden während der Entwicklung die Designentwürfe mittels Beschreibungen, Spezifizierungen und Prototypen gezeigt und von ihnen bewertet, [3]. Dies kann mittels entsprechenden Testumgebungen erfolgen.

Die Arbeit von Nielsen [8] begründet das iterative Vorgehen für die Entwicklung von Designlösungen als hilfreich. Durch Iterationen entstehen verschiedene Versionen und eine Sammlung von Erfahrungen, die in die Entwicklung einfließen. Daraus können Steigerungen der Bedienbarkeit und des Benutzererlebnisses mit dem Design erzielt werden.

Für den Anwendungsfall dieser Arbeit wird nach dem Prozess der ISO 9241–210 vorgegangen und die Gestaltungslösungen mit entsprechenden Testumgebungen bewertet. So

werden anhand der Anforderungen entsprechende Visualisierungen entwickelt und nach jeder Iteration, der Entwurf durch den Austausch mit dem Entwicklerteam verfeinert. Dies kann in einer Testumgebung erfolgen, auf welche schnell zugegriffen und die ohne aufwendige Vorbereitung, ortsunabhängig eingesetzt werden kann.

In Nutzerstudien können entworfene Informationsvisualisierungen gegenübergestellt und die Bedienbarkeit und das Benutzererlebnis bewertet werden. Daraus können Ergebnisse zur Anforderungserfüllung und Eindrücke von Probanden zu den einzelnen Lösungsentwürfen resultieren. Dafür kann wie in Abschnitt 3 eine sichere und realitätsnahe Testumgebung verwendet werden, um Ergebnisse möglichst auf die reale Situation im Fahrzeug übertragen zu können und eine Vielzahl von Bewertungen auf eine einheitliche Situation zu erhalten.

Weitere Nutzerstudien in abschließenden Entwicklungsiterationen können das Gesamtsystem auf die Gültigkeit testen.

4.2 Iterative Entwicklung in der Softwareentwicklung

In der Literatur zur Softwareentwicklung werden agile Prozessmodelle aufgeführt, welche ein iteratives Vorgehen verwenden.

Erickson et al. [5] beschreiben den Begriff *Agil* bei diesen Modellen, dass bei einem Softwareprojekt, die Komplexität so gering wie möglich gehalten wird. So kann schneller auf Änderungen der Umgebung und Nutzeranforderungen reagiert werden. [5]

In der Arbeit von Dybå und Dingsøy [4] werden folgende agile Prozessmodelle aus der Literatur beschrieben und gegenübergestellt.

- Crystal Methodologies
- Dynamic Software Development Method (DSDM)
- Feature-Driven Development
- Lean Software Development
- Scrum
- Extreme Programming (XP, XP2)

Dybå und Dingsøy [4] erwähnen für jedes Modell bestimmte Eigenschaften und Anforderungen, welche Nutzen für diese Arbeit bereitstellen. Allgemein ist aufgefallen, dass ein Softwareprojekt nicht als Gesamtsystem zu einem Zeitpunkt entwickelt werden muss. Wie beispielsweise bei *Scrum* wird das Projekt in kleine Arbeitspakete aufgeteilt und nacheinander von den Entwicklern abgearbeitet. Während der Entwicklung werden Rücksprachen mit involvierten Personen gehalten und darauf aufbauend weiterentwickelt. In *DSDM* wird diese Eigenschaft unter dem Begriff *iterative und inkrementelle Entwicklung* aufgeführt. Es wird zusätzlich darauf hingewiesen, wie in *Extreme Programming*, die Softwareentwicklungen nach jedem Arbeitspaket auf Funktionalität und Anforderungserfüllung zu testen. Jedoch ändern sich auch Anforderungen während der Entwicklung. Aus diesem Grund wird die Software erweiterbar und änderbar aufgebaut. Es ist zusätzlich schwierig, vor unerforschten Entwicklungen den Aufwand einzuschätzen. Diese können durch eine Reflexion der Entwickler nach einer Bearbeitungsphase erfahren werden. Möglicherweise resultieren dadurch Anpassungen am Projektplan, wie es in der *Crystal Methodologies* erwähnt wird. [4]

Softwareentwickler werden in das interdisziplinäre Entwicklungsteam einbezogen und Änderungen sind umgehend umsetzbar. Das ermöglicht die Entwicklung und empirische Untersuchung von Informationsvisualisierungen des automatisierten Fahrens.

Das Testen der Implementierung nach vereinzelten Iterationen ist dabei von essenzieller Bedeutung, welches in entsprechenden Testumgebungen erfolgt. Für einen Softwareentwickler muss die Möglichkeit bestehen, erste Entwicklungen ohne Aufwand und direkt am Arbeitsplatz in einer Testumgebung zu testen.

Zusätzlich sind die gewünschten Anzeigen und Funktionen von AR-Visualisierungen als Erweiterung der physischen Umgebung zu prüfen. Beispielsweise können Farbinformationen von Hintergrund und der Visualisierungen die Wahrnehmung der Anzeige

beeinflussen. Dies kann durch eine Testumgebung erfolgen, in welcher eine reale Umgebung dargestellt wird.

Der Umgang der Software mit realen Daten beeinflusst unter anderem auch die Anzeige der Visualisierungen. Getestet kann dies in einer Testumgebung, in welcher die Daten wie im realen Einsatz verarbeitet werden.

5 Drei Testumgebungen

Auf Basis des gewonnenen Hintergrundwissens und den erarbeiteten Anforderungen werden in diesem Abschnitt drei Testumgebungen vorgestellt, welche im weiteren Projektvorhaben eingesetzt werden, um Visualisierungsentwürfe und Softwareimplementierungen für automatisierte Fahrfunktionen iterativ entwickeln und bewerten zu können. Diese haben charakteristische Eigenschaften und Einsatzmöglichkeiten, welche nachfolgend beschrieben sind.

5.1 Testumgebung mit verkleinertem Modell

Die Testumgebung mit dem verkleinerten Modell repräsentiert eine vereinfachte Abbildung einer realen Straßenkreuzung, in der Visualisierungsentwürfe angezeigt werden können. Der Aufbau und die Eigenschaften dieser Testumgebung entstand aus den beschriebenen Herausforderungen der ersten Entwicklungsphase, bei welcher die Testumgebung im realen Fahrzeug eingesetzt worden ist. Es sollte eine Testumgebung bei der Entwicklung von neuen Visualisierungsentwürfen zur Verfügung stehen, bei welcher

- die Vorbereitung und der Zugriff schnell und ohne großen Aufwand erfolgen kann,
- diese ortsunabhängig eingesetzt werden kann und
- Verfeinerungs- und Änderungshinweise unkompliziert eingearbeitet werden können.

Das verkleinerte Modell, siehe Abbildung 3, ist eine Karte einer realen Straßenkreuzung im Maßstab 1:167. Diese ist in der Draufsicht

ausgedruckt. Zur Verdeutlichung der Unübersichtlichkeit werden Styropor-Würfel, welche Häuser repräsentieren, auf die Karte geklebt. Einen Kontext für bestimmte Visualisierungen schaffen erstellte Simulationen der Straßenverkehrssituation. Beispiele hierfür sind virtuelle Fahrzeuge, welche über die Karte fahren.



Abbildung 3: Das verkleinerte Modell

Als Vorbereitung muss die Karte einmalig, wie zuvor beschrieben, aufgebaut und jeweils die Informationsvisualisierung implementiert werden. Diese müssen zur Darstellung auf den Maßstab angepasst und der virtuellen Miniaturkarte hinzugefügt werden. Die physische Karte kann am Arbeitsplatz vom Entwicklerteam aufgebaut und mit Visualisierungen erweitert werden. Somit kann auf die Entwicklungen schnell zugegriffen und diese mit dem Anzeigegerät getestet werden.

Zur genauen Anzeige der Visualisierungen müssen deren Positionen mit der Karte übereinstimmen. Dafür ist die Funktionalität des Marker-Trackings zur Positionsbestimmung, als Referenz der Koordinaten von Visualisierungen, hinzugefügt worden. Diese Funktion ist aus den Arbeiten übernommen worden, welche in Abschnitt 3 beschrieben sind. Somit kann die Testumgebung an jedem Ort und unabhängig der Raumgröße eingesetzt werden, was hinsichtlich der Vorbereitung und Durchführung, im Vergleich zum realen Fahrzeug, einen deutlich geringeren Aufwand darstellt. Beispielsweise kann das Modell vom Arbeitsplatz in Besprechungsräume getragen werden, um mit involvierten Personen des Projekts Hinweise zu Verfeinerungen der Lösung auszudiskutieren. Der Wechsel vom Entwicklungs- zum Diskussionsort ermöglicht unter

anderem schneller Änderungshinweise einzu- arbeiten als mit der Testumgebung im realen Fahrzeug. Aufgrund des geringen Vorbereitungs- aufwands können öfters Besprechungen am Modell durchgeführt werden.

Zusammengefasst eignet sich diese Testum- gebung als Austauschmedium im Entwicklungs- team. Jedoch wird angenommen, dass sich diese Testumgebung nicht dafür eignet, Nut- zerbewertungen der entsprechenden Visuali- sierungsentwürfe zu erfahren. Begründet wird diese Aussage damit, dass sich die Umgebung zu sehr der im realen Fahrzeug unterscheidet.

Für einen internen Technologietag der Robert Bosch GmbH zum Thema *Human-Machine Interaction* konnte das Modell mit den proto- typisch umgesetzten Visualisierungen benutzt werden, um das Projektvorhaben vorzustellen. Die Karte ist ausgedruckt und auf einen erhöh- ten Tisch geklebt worden. Ein entworfenes Szenario zum Verständnis des Einsatzes als Kommunikationsmedium ist nachfolgend beschrieben.

Eine Person setzt die HoloLens auf und schaut in Fahrgastperspektive in die Karte. Zu Beginn werden keine Visualisierungen mit Informati- onen des Connected Automated Driving ange- zeigt. Es sind ausschließlich fahrende, virtuelle Fahrzeuge realitätsnah mit Verdeckungen auf der Straße zu sehen. Die Situation beschreibt, ohne abzubremesen in eine Vorfahrtstraße einzufahren. Anschließend werden die Informationsvisualisierungen ein- geblendet. Beispielsweise Visualisierungen zum Hinweis auf Fahrzeuge, welche von Häu- sern verdeckt sind.

Das Feedback der Personen aus verschiedenen Fachrichtungen bestätigt den Nutzen, Informati- onen des automatisierten Fahrens anzuzei- gen, um Fahrentscheidungen nachzuvollziehen. Die Testumgebung eignet sich somit auch, eine Projektidee vorzustellen.

5.2 Testumgebung im Fahrsimulator

Eine realitätsnahe Fahrumgebung in einem Straßenabschnitt stellt die Testumgebung im

Fahrsimulator bereit, mit welcher Visualisie- rungsentwürfe und -entwicklungen getestet werden können. Diese Testumgebung entstand zur Lösung von Herausforderungen, die wäh- rend der ersten Entwicklungsphase mit der Testumgebung im realen Fahrzeug aufgefallen sind. Diese Testumgebung erfüllt folgende Kriterien:

- Es soll die Möglichkeit bestehen, bestimmte Verkehrssituationen wie- derholt darzustellen.
- Aussagen und Ergebnisse sollen auf den Einsatz im realen Fahrzeug übertragen werden können.

Mit der Testumgebung im Fahrsimulator besteht die Möglichkeit, Visualisierungen in realer Größe vor einer zweidimensionalen Abbildung einer Straßensituation anzuzeigen. In dieser können projektinvolvierte Personen und potenzielle Nutzer die Entwicklungen bewerten. Das Entwicklerteam kann die Anzeige und die Aussage der AR-Entwicklun- gen in der physischen Umgebung betrachten und Anmerkungen zum gewünschten Lösungsentwurf treffen, welche in weiteren Iterationen einzuarbeiten sind. In Nutzerstu- dien können Aussagen und Bewertungen zu neuen Entwicklungen erfahren werden. Das Einbeziehen dieser Personen in die Entwick- lung ist eine Anforderung des iterativen Ent- wicklungsvorgehens, welches in Abschnitt 4 beschrieben ist.

Als Vorbereitung der Testumgebung im Fahr- simulator müssen Aufbauten vorbereitet werden, welche in Abbildung 4 zu sehen sind.



Abbildung 4: Testumgebung im Fahrsimulator

Es ist ein großes zweidimensionales Anzeigemedium erforderlich, welches *Powerwall* genannt wird. Diese ist eine Wand aus einem halbdurchlässigen Material, auf welche ein Bild von Projektoren hinter der Wand abgebildet wird. Vor diese Wand ist ein sogenanntes *Car-Mockup* positioniert, welches eine durchgesägte Fahrzeugkarosse mit Beifahrersitz darstellt. Damit wird dem Fahrgast eine dem realen Fahrzeug ähnelnde Umgebung geschaffen, wie es in Abschnitt 3 beschrieben ist.

Mit der Powerwall werden aufgenommene Videos einer bestimmten Verkehrssituation abgespielt. Die Sicht auf den Verkehr der Probanden im Fahrzeug wird mittels der HoloLens-Anzeige von Visualisierungen erweitert. Die Testumgebung ist so aufgebaut, dass in Tests zahlreiche Personen die Entwicklungen mit identischen Verkehrssituationen erfahren können.

Einblendungen von Informationsvisualisierungen erfolgen im konkreten Anwendungsfall meist im fahrenden Zustand des Fahrzeugs. Das AR-Erlebnis muss dem Nutzer dabei so bereitgestellt werden, als würde er sich tatsächlich in der realen Umgebung befinden. Die zu lösende Herausforderung ist, die stetige Anpassung von Positionswerten von Visualisierungen in Bezug auf die sich dynamisch verändernde 2D-Abbildung der physischen Umgebung in einem Video. Für die erste Vorbereitung der Testumgebung ist eine sogenannte *Keyframing-Methode* zur Anpassung von Positionswerten als Unterstützung der Softwareentwicklung implementiert worden. Dabei können einzelne Visualisierungen nach Einschätzungen des Entwicklers, passend auf einen bestimmten Zeitpunkt im Video, transformiert werden. Zwischen den Positionswerten wird dann beim Abspielen der Simulation interpoliert.

Die Ergebnisse des realitätsnahen Testens sind, wie in Abschnitt 3 beschrieben, nicht vollständig auf die Realität übertragbar, da in dieser Umgebung nicht alle Einflüsse des realen Fahrzeugs implementiert werden können. Diese Testumgebung dient für Voruntersuchungen und für Tests in Situationen, die aus sicherheitskritischen Gründen nicht im

realen Fahrzeug durchzuführen sind. Personen können in einer sicheren Umgebung die Applikation auf die Bedienbarkeit, das Benutzererlebnis und Aussage der Visualisierungen testen. Die Vorbereitung ist aufwendig, jedoch kann aus den Ergebnissen der Tests profitiert werden, da diese für jeden Probanden identisch und somit reproduzierbar sind.

Für einen realen Nutzertest der Robert Bosch GmbH ist diese Testumgebung vorbereitet und eingesetzt worden. Dies bestätigt die Annahme, dass mit dieser Testumgebung Tests mit potenziellen Nutzern durchgeführt werden können. Es können damit Eindrücke und Bewertungen in die Entwicklung von AR-Visualisierungen im Anwendungsfall des automatisierten Fahrens einfließen.

5.3 Testumgebung im realen Fahrzeug

Diese Testumgebung repräsentiert das Zielsystem und ist, wie in Abschnitt 1 erwähnt, in der beginnenden Projektphase von anderen Personen des Projektteams entwickelt und eingesetzt worden. Dabei sind die genannten Herausforderungen aufgetreten, welche mit den zuvor erwähnten Testumgebungen gelöst worden sind. Mit dieser Testumgebung können Ergebnisse erzielt werden, welche aus simulierten Umgebungen nicht zu erfahren sind. Diese Testumgebung im realen Fahrzeug erfüllt folgendes Kriterium:

Entwickelte Visualisierungen sollen mithilfe vom Fahrzeug bereitgestellten Daten, in der realen physischen Umgebung angezeigt und getestet werden können.

Diese Testumgebung repräsentiert das Prüfen von entwickelten Designentwürfen, welche in vorangehenden Iterationen verfeinert worden sind. Anders als in den vorangehenden Tests, kann im Umfeld des Endsystems und mit realen Daten getestet werden. Das Entwicklungsteam kann sich über die Entwicklung am Standort austauschen und es können Nutzerstudien durchgeführt werden.

Anders als die zuvor erwähnten Testumgebungen werden für diese Variante keine Simulationen, sondern reale Daten aus dem Connected

Automated Driving verwendet, um den Visualisierungen entsprechende Eigenschaften und Verhalten zu geben. Die von der Infrastruktur erfassten Daten werden dem Fahrzeug über Mobilfunknetze in Echtzeit übertragen. Sensoren sind dafür an einem Straßenabschnitt in Ulm aufgebaut worden, um diesen zu überwachen und dem System Informationen zu detektierten Fahrzeugen zu liefern. Dieser Straßenabschnitt beschreibt eine Pilotanlage für das Projekt *Mobile Edge Computing basierte Objekterkennung für hoch- und vollautomatisiertes Fahren (MEC-View)*, welches den Kontext dieser Arbeit bildet. Für diese Testumgebung wird eine Netzwerkverbindung zwischen dem Fahrzeugcomputer und der HoloLens gefordert. Die Daten werden auf dem *Robot Operating System (ROS)* des Fahrzeugs in ein Datenformat serialisiert. Das sogenannte *Protocol buffers* von Google wird als Datenformat zur Serialisierung verwendet. Die verarbeiteten Daten werden an die Applikation der HoloLens übertragen und dort deserialisiert. Zur Anzeige mit AR in die Sicht der realen Umgebung, werden die Daten in Echtzeit zur Verfügung gestellt.

Diese Testumgebung ist sehr aufwendig in der Vorbereitung und bietet keine Garantie, dass Ergebnisse mit entsprechenden Tests erzielt werden können. Es muss vom Entwicklungsstandort zu einer Pilotanlage nach Ulm gefahren und die entsprechenden Tests dort durchgeführt werden. Es existieren Informationen, welche von bestimmten Situationen ausgehen. Diese sind nicht immer reproduzierbar, da so die Sicherheit des realen Straßenabschnitts nicht gewährleistet werden kann.

Für die erste Entwurfsphase bietet diese Testumgebung keinen Nutzen, jedoch gliedert sich diese in das Projektvorgehen als letzte Phase, um Aussagen des Gesamtsystems zu erhalten, die mit den anderen nicht getroffen werden können.

6 Zusammenfassung und Ausblick

In dieser Arbeit sind drei erarbeitete Testumgebungen beschrieben, die in einem iterativen Entwicklungsvorgehen Einsatz finden können.

Damit ist ein wichtiges Thema betrachtet worden, welches die zukünftige Entwurfsentwicklungen von gültigen AR-Informationsvisualisierungen, zur Anzeige während des automatisierten Fahrens, unterstützt. Jedoch können die Erkenntnisse, nach Meinung des Autors, auch in anderen Anwendungsentwicklungen Einfluss finden, um AR-Visualisierungen für verschiedenen Fahrsituationen zu entwickeln.

Untersuchungen von wissenschaftlichen Arbeiten und Einsätze in realen Tests liefern Erkenntnisse, welche in diese Arbeit eingeflossen sind. Lösungsentwürfe können mit den vorgeschlagenen Testumgebungen erarbeitet und entwicklungsbegleitend von verschiedenen Personen in den Testumgebungen bewertet werden.

Das Projektteam aus verschiedenen Fachrichtungen kann sich orts- und situationsunabhängig über Entwicklungen austauschen und Änderungsempfehlungen vorschlagen. Ermöglicht werden kann damit eine häufige Wiederholung von Tests und eine schnelle Einarbeitung von Verbesserungshinweisen. Zu diesen Testumgebungen zählen die mit einem verkleinerten Modell und in einem Fahrsimulator. Das verkleinerte Modell kann mit einem geringen Aufwand der Vorbereitung genutzt werden, um sich schnell und ortsunabhängig über erste Entwicklungserfolge austauschen zu können. Im Fahrsimulator wird eine realitätsnahe Umgebung geschaffen, sodass sich Bewertungen möglichst auf den realen Einsatz im Fahrzeug übertragen lassen, was mit dem Miniaturmodell nicht erzieht werden kann. Zu diesen Bewertungen gehören Nutzerstudien, welche Ergebnisse zu den Visualisierungsentwürfen aus Sicht von potenziellen Nutzern liefern. Die Testumgebung im realen Fahrzeug gliedert sich in die Entwicklung, um reale Einflüsse während der Fahrt in einem autonomen Fahrzeug in die Bewertung einfließen zu lassen. Diese Einflüsse können in den statischen Umgebungen nicht simuliert werden.

Jede Testumgebung repräsentiert eigene Eigenschaften und hat eine eigene Vorbereitung. Zukünftig kann der Vorbereitungsaufwand jedoch durch Übertragbarkeiten von

Systemfunktionen und -komponenten weiter reduziert werden. Beispielsweise sind die erstellten Simulationen zum Zeitpunkt der Bearbeitung dieser Arbeit nicht auf andere Testumgebungen übertragbar und müssen einzeln für die jeweiligen separat erstellt werden. Neben der Reduzierung des Vorbereitungsaufwandes, können auch Unklarheiten von Personen während einer Präsentation gelöst werden, indem zur Erläuterung in eine simulierte Testumgebung gewechselt wird. Diese Funktionssimplementierungen übernehmen zukünftige Weiterentwicklungen.

7 Literatur

- [1] R. T. Azuma, „A survey of augmented reality“, *Presence: Teleoperators & Virtual Environments*, Bd. 6, p. 355–385, 1997.
- [2] M. Dias, J. Jorge, J. Carvalho, P. Santos und J. Luzio, „Usability evaluation of tangible user interfaces for augmented reality.“ in *2003 IEEE International Augmented Reality Toolkit Workshop*, 2003.
- [3] DIN EN ISO 9241-210:2011-01, „Ergonomie der Mensch-System-Interaktion - Teil 210: Prozess zur Gestaltung gebrauchstauglicher interaktiver Systeme (ISO 9241-210:2010); Deutsche Fassung EN ISO 9241-210:2010“, DIN Deutsches Institut für Normung e. V., 2011.
- [4] T. Dybå und T. Dingsoyr, „Empirical studies of agile software development: A systematic review“, *Information and software technology*, Bd. 50, p. 833–859, 2008.
- [5] J. Erickson, K. Lyytinen und K. Siau, „Agile modeling, agile software development, and extreme programming: the state of research“, *Journal of Database Management (JDM)*, Bd. 16, p. 88–100, 2005.
- [6] P. Milgram und F. Kishino, „A taxonomy of mixed reality visual displays“, *IEICE TRANSACTIONS on Information and Systems*, Bd. 77, p. 1321–1329, 1994.
- [7] A. Nawab, K. Chintamani, D. Ellis, G. Auner und A. Pandya, „Joystick mapped augmented reality cues for end-effector controlled tele-operated robots“, in *2007 IEEE Virtual Reality Conference*, 2007.
- [8] J. Nielsen, „Iterative user-interface design“, *Computer*, Bd. 26, p. 32–41, 1993.
- [9] L. Pfannmüller, „Anzeigekonzepte für ein kontaktanaloges Head-up Display“, München, 2017.
- [10] F. Sauer, F. Wenzel, S. Vogt, Y. Tao, Y. Genc und A. Bani-Hashemi, „Augmented workspace: Designing an AR testbed“, in *Proceedings IEEE and ACM International Symposium on Augmented Reality (ISAR 2000)*, 2000.
- [11] D. W. F. Van Krevelen und R. Poelman, „A survey of augmented reality technologies, applications and limitations“, *International journal of virtual reality*, Bd. 9, p. 1–20, 2010.
- [12] M. Walch, J. Frommel, K. Rogers, F. Schüssel, P. Hock, D. Dobbstein und M. Weber, „Evaluating VR driving simulation from a player experience perspective“, in *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, 2017.

Entwicklung eines Prototypen zur Müdigkeitserkennung im Fahrsimulator

Sven Vennebusch

Hochschule Reutlingen

sven_hendrik.vennebusch@Student.Reutlingen-University.de

Abstract

Ein nicht unerheblicher Anteil der Autounfälle ist auf Müdigkeit am Steuer zurückzuführen. Um Unfälle aufgrund von Müdigkeit zu vermeiden, existieren schon einige Ansätze wie beispielsweise die Erkennung der Fahrweise. Im Rahmen des IOT-Labors des Masterstudiengangs Human Centered Computing der Hochschule Reutlingen sollen verschiedene Fahrassistenzsysteme entwickelt und getestet werden, um Unfälle aufgrund von Müdigkeit zu verhindern. Diese Arbeit beschäftigt sich mit der Müdigkeitserkennung über Computer Vision (CV) und das Elektrokardiogramm (EKG). Im Rahmen dieses Papers wird die Müdigkeitserkennung über CV am Steuer mittels den Open Source Bibliotheken OpenCV und Dlib und dem Embedded PC Nvidia Jetson Nano verwirklicht. Die Müdigkeit über EKG wird über den Herzschlag und die Herzfrequenzvariabilität erkannt. Ebenfalls wurde in dieser Arbeit eine Schnittstelle aus CV und EKG entwickelt, um aus den Python-Skripten der Müdigkeitserkennung über Computer Vision und der Müdigkeitserkennung über EKG die

zur Erkennung wichtigen Daten zusammenzufassen. Diese werden anschließend zu einem gesamten Ergebnis ausgewertet.

Schlüsselwörter

Computer Vision, Dlib, Elektrokardiogramm, Fahrassistenzsysteme, Müdigkeitserkennung, OpenCV, Python

CR-Kategorien

- Software and its engineering ~ Software creation and management
- Applied computing ~ Life and medical sciences

1 Einleitung

Immer wieder kommt es zu Unfällen aufgrund von Müdigkeit am Steuer. Laut ADAC gaben 26% der befragten Personen an, schon einmal am Steuer eingeschlafen zu sein.¹ Dabei kann bereits ein Sekundenschlaf von einer Sekunde fatale Folgen haben. Bei einer Geschwindigkeit von 130 km/h legt man bereits 36 Meter im Blindflug zurück. Schätzungsweise 2000 Unfälle pro Jahr, bei denen Personen zu Schaden kommen, werden durch das Einschlafen des Fahrers verursacht. Die Dunkelziffer wird wohl weit höher liegen.

Auf Autobahnen zählt jeder vierte tödliche Unfall mit Pkw dazu. Kommt es zu einem Unfall aufgrund von Müdigkeit, muss der Fahrer laut Strafgesetzbuch § 315 c mit einer

Betreuer Hochschule: Prof. Natividad Martinez
Hochschule Reutlingen
natividad.martinez@Reutlingen-University.de

Wissenschaftliche Vertiefungskonferenz
20. Mai 2020, Hochschule Reutlingen
Copyright 2020 Sven Vennebusch

¹ <https://www.adac.de/verkehr/verkehrsmedizin/muedigkeit-sekundenschlaf-auto/>, Zugriff: 13.03.20

Geldstrafe oder fünf Jahren Gefängnis rechnen. Hinzu kommt in jedem Fall der Fahrerlaubnisentzug ².

Aus diesem Grund wurde das deutsch-ägyptische Forschungsprojekt MuFaDe ins Leben gerufen. In diesem soll untersucht werden, wie der Fahrer vor Müdigkeit gewarnt werden kann.

Wichtige Messgrößen zur Erkennung der Müdigkeit sind die Atmung, der Puls, die Augenbewegung (blinzeln) und die Gähnrate. Im Automobilbereich spiegeln sich zwei generelle Richtungen für die Müdigkeitserkennung wider: die Müdigkeitserkennung über EKG (Elektrokardiogramm) bzw. Herzrate [1,2] und die Müdigkeitserkennung über Bildverarbeitung (Computer Vision) [3,4].

An der Hochschule Reutlingen wurde bereits die Müdigkeitserkennung über EEG (Elektroenzephalografie) in einer Masterarbeit validiert. Ebenfalls wurden im Rahmen des Masterprojekts IOT von M. Gromer und D. Salb die Müdigkeitserkennung über EKG [5] und von S. Vennebusch [6] die Müdigkeitserkennung über CV erarbeitet. Beide Projekte wurden auf dem Embedded PC Nvidia Jetson TX2 entwickelt. Da dieser sehr kostenintensiv ist, soll der Prototyp für den Fahrsimulator auf dem Nvidia Jetson Nano laufen, welcher nur einen Bruchteil des Jetson TX2 kostet. Im Rahmen dieses Projektes soll die Müdigkeitserkennung über Bildverarbeitung und Elektrokardiogramm auf dem Nvidia Jetson Nano implementiert werden, um so einen Prototyp für den Fahrsimulator der Hochschule Reutlingen zur Verfügung zu stellen. Hierfür soll ebenfalls ein einfacher prototypischer Code für die endgültige Müdigkeitserkennung implementiert werden, welcher die Erkennung über Computer Vision und Elektrokardiogramm zusammenfasst.

1.1 Problemstellung und Ziele der Arbeit

Diese Arbeit befasst sich mit der Entwicklung eines Prototyps zur Müdigkeitserkennung im Fahrsimulator mittels Computer Vision und Elektrokardiogramm. Die Grundsteine hierfür wurden bereits in [5] und [6] gelegt. Da diese allerdings auf dem kostenintensiven Nvidia Jetson TX2 entwickelt wurden, müssen diese auf dem kostengünstigeren Nvidia Jetson Nano implementiert werden. Hierbei kommt es zu Änderungen der Hardware und zu kleinen Änderungen der Software. Diese sollen in dieser Arbeit umgesetzt werden. Ebenfalls soll ein einfacher Algorithmus entwickelt werden, um ein endgültiges Ergebnis der Müdigkeit aus Computer Vision und Elektrokardiogramm zu erhalten. Dieser Algorithmus soll allerdings nur prototypisch sein, um mit diesem durch Tests Evaluierungen zur Müdigkeitserkennung durchführen zu können.

2 Physiologische Grundlagen

In diesem Kapitel sollen die physiologischen Grundlagen geschaffen werden, wie Müdigkeit mittels Gesichtszüge (Computer Vision) und Herzschlag bzw. Herzfrequenzvariabilität (Elektrokardiogramm) erkannt werden kann. Als erstes soll die Müdigkeitserkennung über Gesichtszüge erklärt werden.

Im darauffolgenden Kapitel wird die Müdigkeitserkennung über das Elektrokardiogramm vorgestellt. Hierfür soll als erstes erläutert werden, wie ein Elektrokardiogramm funktioniert und was ein QRS-Komplex ist. Anschließend werden die Erkennung über den Herzschlag sowie zwei mögliche Erkennungen über die Herzfrequenzvariabilität erklärt.

² <https://www.bussgeldkatalog.org/muedigkeit-autofahren>, Zugriff: 12.03.20

2.1 Gesichtszüge

Ziel der Müdigkeiterkennung über Computer Vision ist eine möglichst objektive Einschätzung der Müdigkeit des Fahrers. Dies wird durch die Verwendung von standardisierten Skalen gewährleistet.

Zwei der am häufigsten verwendeten Skalen sind die Skala von Wierwille und Ellsworth [7] sowie die Skala von der Human Factors Consult GmbH Berlin (HFC) [8]. Die Skala nach Wierwille und Ellsworth umfasst vier Müdigkeitsstufen (von 1 = nicht müde bis 4 = Sekundenschlaf). Diese werden durch Verhaltens- und Augenparameter beschrieben. Die Skala von HFC umfasst neun Müdigkeitsstufen (1 = wach bis 9 = Schlaf).

In den folgenden beiden Abschnitten wird die Skala nach Wierwille und Ellsworth kurz vorgestellt, anschließend die Begründung für die Wahl der verwendeten Skala erklärt.

2.1.1 Müdigkeitsbestimmung nach Wierwille und Ellsworth

Die Skala von Wierwille und Ellsworth (auch Wierwille-Skala genannt) differenziert die Müdigkeit eines Fahrers in vier Stufen. Stufe 1 bedeutet nicht müde, Stufe 4 bedeutet Sekundenschlaf. Unterschieden werden die Stufen anhand der Augenparameter und der Verhaltensindikatoren. Während den wachen Fahrer (Stufe 1) häufige und schnelle Blicke auszeichnet, nehmen diese bis Stufe 3 stark ab und enden in einem starren Blick. Bereits ab einer Lidschlagdauer von über 500ms wird von einem Mikroschlaf gesprochen. Ebenfalls nimmt im Laufe der Ermüdung die Lidschlaghäufigkeit ab, während sich die Dauer des Lidschlags verlängert (Stufe 3). Beim Sekundenschlaf (Stufe 4) treten Lidschläge über einer Sekunde auf. Ebenfalls Indikatoren für die Müdigkeit des Fahrers sind nach Wierwille und Ellsworth eine zunehmend bequemere Sitzhaltung und eine erschaffende Mimik. Weiterhin können beim Sekundenschlaf Schreckreaktionen beobachtet werden, welche durch das schreckhafte Aufwachen ausgelöst werden. Diese können

ebenfalls an einer überschießenden Lenkbewegung zu erkennen sein. Die vier Müdigkeitsstufen in Bezug auf Augenparameter nach Wierwille und Ellsworth sind in Tabelle 1 zusammengefasst.

Tabelle 1: Müdigkeitsstufen nach Wierwille und Ellsworth anhand der Augenparameter

Müdigkeitsstufe	Beschreibung
Stufe 1 (Wach)	Kurze und häufige Lidschläge, schnelle Blicke
Stufe 2 (Leichte Müdigkeit)	Kurze, aber weniger häufige Lidschläge, langsame Blicke
Stufe 3 (Starke Müdigkeit)	Lange Lidschläge (über 500ms), starrer Blick
Stufe 4 (Sekundenschlaf)	Sehr lange Lidschläge (über einer Sekunde)

2.1.2 Verwendete Skala

Für den Prototypen wird die Wierwille-Skala verwendet, da diese die Müdigkeit in weniger Stufen unterteilt und einen Ansatz bietet, die Müdigkeit in Echtzeit zu bewerten im Gegensatz zur Skala von HFC, welche nur die Möglichkeit bietet, im Nachhinein die Müdigkeit des Fahrers zu bewerten.

Zur Müdigkeiterkennung über Gähnen gibt es bisher keine festen Skalen. Die Dauer eines durchschnittlichen Gähnvorgangs beträgt ungefähr 3,5 bis 6 Sekunden, zeigt aber eine große Variation. Meist wird mehrere Male hintereinander gegähnt, wobei die Intervalllänge bei etwa einer Minute liegt. Zwischen der Häufigkeit und der Dauer des Vorgangs gibt es keinen signifikanten Zusammenhang [9].

Als Ergebnis des Projektes soll Müdigkeit anhand der Augenparameter nach Stufe 3 der Wierwille-Skala erkannt werden sowie Gähnen.

Im Laufe der Arbeit müssen folglich 4 Schwellenwerte definiert werden. Zum einen die Werte, ab wann der Mund und das Auge als geschlossen gelten, zum anderen, ab welcher Zeit der Fahrer als müde einzustufen ist.

2.2 Elektrokardiogramm

In diesem Abschnitt soll die Müdigkeitserkennung mit Hilfe des Elektrokardiogramms erklärt werden. Als erstes wird hierfür erklärt, was ein Elektrokardiogramm ist und was mit diesem gemessen werden kann. Anschließend werden die Möglichkeiten erklärt, wie man mit dem EKG Müdigkeit erkennen kann.

2.2.1 Funktionsweise

An der rechten Vorhofwand des Herzens befindet sich der Sinusknoten. Er sendet elektrische Impulse, die sich über das Herz ausbreiten, damit es gleichmäßig schlägt. Beim Elektrokardiogramm messen Elektroden, wie sich die Herzströme im Herzen ausbreiten und zurückbilden. Die Elektroden sind an Brustkorb, Armen und Beinen befestigt und erfassen die elektrische Erregung an der Hautoberfläche. Abbildung 1 zeigt, wie diese elektrischen Reize am Körper abgegriffen werden können.

Zum Messen dieser elektrischen Reize kommen nach der klassischen Einthoven-Ableitung drei Elektroden zum Einsatz. Diese drei Elektroden bilden das sogenannte Einthoven-Dreieck um das Herz. Oft werden die drei Elektroden an beiden Armen sowie am linken Bein befestigt. Allerdings ist es sinnvoller, die Elektroden in der Nähe des Herzens zu befestigen. Dabei eignen sich vor allem die Positionen zwischen der rechten bzw. linken Schulter und der rechten bzw. linken Brust sowie links unterhalb der Rippen.

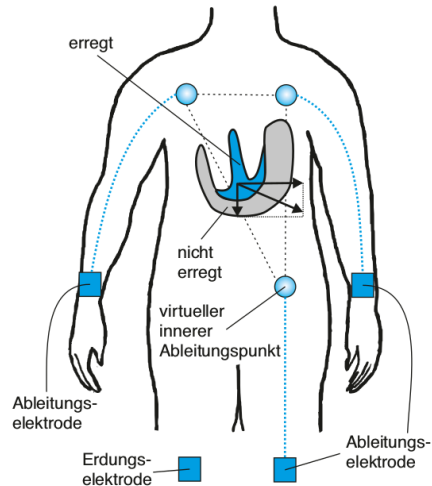


Abbildung 1: Reizleitung und Anschluss der Elektroden [10]

In Abbildung 2 ist das Einthoven-Dreieck mit den sich daraus ergebenden drei Ableitungen der elektrischen Reize zu sehen.

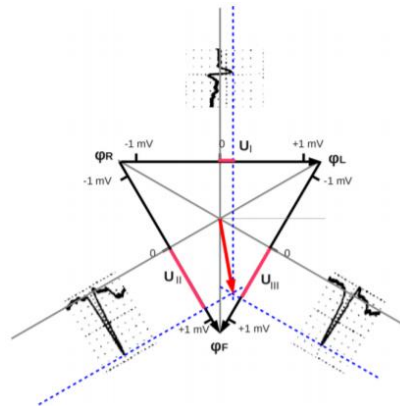


Abbildung 2: Das Einthoven-Dreieck mit seinen drei Ableitungen³

Der Punkt links oben ist der rechte Arm, der Punkt rechts oben der linke Arm und der untere Punkt das linke Bein. Der rot dargestellte

³ http://physiologie.medizin.uni-leipzig.de/files/EKG_2_8_Pharma.pdf, Zugriff: 12.03.20

Vektor ist der sogenannte Summationsvektor. Dieser stellt die Richtung des elektrischen Reizes und somit auch die Orientierung des Herzmuskels dar. In Abbildung 2 ist ebenfalls zu sehen, dass die größte Spannung entlang der Herzachse zwischen dem rechten Arm und dem linken Bein gemessen werden kann. Werden die so gemessenen Spannungen über die Zeit aufgetragen, entsteht ein Elektrokardiogramm. In Abbildung 3 ist das Elektrokardiogramm im zeitlichen Verlauf zu sehen. Gut zu sehen ist hierbei der sogenannte QRS-Komplex. Der QRS-Komplex ist ein Kurvenbestandteil des Elektrokardiogramms. Er besteht normalerweise aus drei Elementen:

- einer negativen Q-Zacke
- einer positiven R-Zacke
- einer kleinen negativen S-Zacke

Dieser zeigt die Erregungsausbreitung im Herzen.

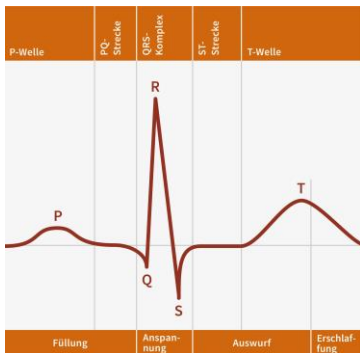


Abbildung 3: QRS-Komplex im EKG ⁴

Der QRS-Komplex steht im übertragenen Sinne für einen Herzschlag. Dieser kann gemessen werden und so die Herzrate bestimmt werden.

2.2.2 Müdigkeitserkennung mittels Elektrokardiogramm

Die Müdigkeitserkennung mittels EKG kann mit Hilfe der Herzrate und mit Hilfe der Herzfrequenzvariabilität festgestellt werden. Für den Prototypen werden beide Varianten verwendet.

2.2.2.1 Herzrate

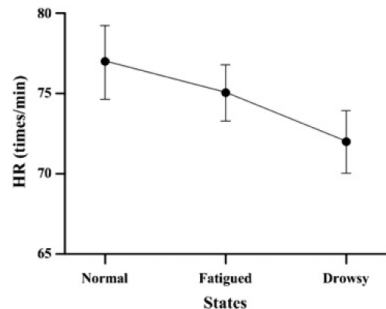


Abbildung 4: Verlauf der Herzrate nach [11]

Eine Möglichkeit, mittels EKG Müdigkeit zu erkennen, ist die Herzrate. Die Herzrate beschreibt die Schläge pro Minute des Herzens. Verschiedene wissenschaftliche Quellen bestätigen unabhängig voneinander, dass mit zunehmender Müdigkeit die Herzrate sinkt und diese somit ein Indikator für Müdigkeit ist [11][12][13]. Zu unterscheiden sind ein Stadium der normalen Müdigkeit und das Stadium einer Müdigkeit, in welcher der Fahrer schon kurz vor dem Einschlafen ist. Das zweite Stadium ist für eine Müdigkeitserkennung im Fahrsimulator ungeeignet, da bei diesem Stadium der Fahrer bereits übermüdet ist. [11] und [13] sind sich einig, dass das erste Stadium der Müdigkeit mit einer Herzrate zwischen 70 und 75 erreicht ist. Der übermüdete Fahrer hat nach [11] und [12] eine Herzrate von ca. 58 bis 64. In Abbildung 4 ist der Verlauf der Herzrate für das erste

⁴ <https://www.ratgeber-herzinsuffizienz.de/de/erkennen/herzinsuffizienz-diagnose/ekg>, Zugriff: 14.03.20

Stadium der Müdigkeit nach [11] zu sehen („Drowsy“).

2.2.2.2 Herzfrequenzvariabilität

Eine weitere Möglichkeit, die Müdigkeit eines Fahrers festzustellen, ist die Herzfrequenzvariabilität (engl. *heart rate variability*, HRV). Die Herzfrequenzvariabilität bezeichnet die Fähigkeit des menschlichen Organismus, die Frequenz des Herzrhythmus zu verändern. Ein gesunder Organismus passt die Herzschlagrate beständig den momentanen Erfordernissen an. Eine körperliche Beanspruchung oder auch eine psychische Belastung hat in der Regel eine Erhöhung der Herzfrequenz zur Folge. Diese geht bei Entlastung oder Entspannung wieder zurück. Die Herzfrequenzvariabilität gibt somit die Schwankung der Herzfrequenz bzw. der Herzrate an.

Zum Messen der Herzfrequenzvariabilität gibt es zwei verschiedene Verfahren. Eine einfache statistische Größe ist die Berechnung der Standardabweichung (SDNN) der RR-Intervalle. Ein RR-Intervall ist die Zeit zwischen zwei R-Zacken im EKG (siehe Abbildung 3). Mit zunehmender Müdigkeit sinkt die Standardabweichung der RR-Intervalle [13] [14]. Dieses Verfahren wird auch aufgrund der vorausgegangenen Arbeit [5] an der Hochschule Reutlingen für den Prototyp verwendet.

Ein weiteres Verfahren ist die Spektralanalyse. Diese ist ein sehr genaues Verfahren zur Feststellung der Frequenzanteile, aus denen sich die Variabilität der Herzfrequenz zusammensetzt. Diese gibt beispielsweise Auskunft über die Kopplung von Atmung und Herzschlag im entspannten Zustand. Sind Atmung und Herzschlag gut gekoppelt, ergibt sich in der Spektralanalyse ein klarer Peak. Die Messung wird dabei in drei Frequenzbänder aufgeteilt, VLF (*very low frequency*), LF (*low frequency*) und HF (*high frequency*). Für die Müdigkeitserkennung des Fahrers werden die Frequenzbänder LF und HF verwendet.

Das parasympathische Nervensystem zielt auf den Zustand der Ruhe und Erholung. Es spiegelt sich im HF-Frequenzband wider. Mit zunehmender Ruhe steigt die Frequenz des HF-Frequenzbandes [14].

Das sympathische Nervensystem zielt auf einen auf Aktivität abzielenden Zustand. Es spiegelt sich im LF-Frequenzband wider. Somit wird bei sinkender Aktivität die Frequenz des LF-Frequenzbandes niedriger [14]. In Abbildung 5 ist der Verlauf von LF und HF bei einer Dauer von 120min zu sehen. Getestet wurden zehn Probanden. Diese mussten für 120 min im Fahrsimulator fahren und wurden zunehmend müder.

Ein guter Indikator für Müdigkeit am Steuer ist somit das LF/HF-Verhältnis. Mit zunehmender Müdigkeit sinkt das Verhältnis.

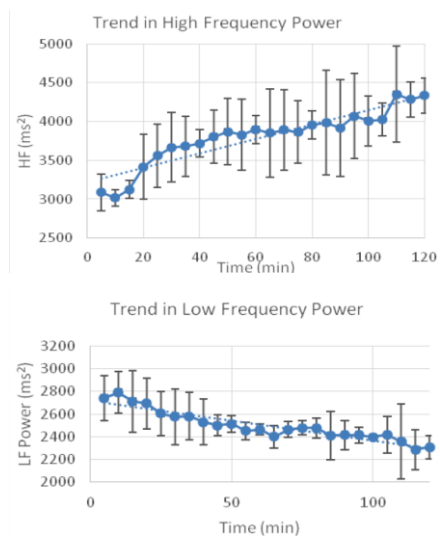


Abbildung 5: Diagramm von Durchschnittswerten von 10 Probanden, welche 120min im Fahrsimulator gefahren sind [14].

3 Der Prototyp

In diesem Kapitel soll der entwickelte Prototyp vorgestellt werden. Hierzu wird als erstes

kurz die verwendete Hardware für das System beschrieben, anschließend die verwendeten Bibliotheken und die Funktionsweise der Software. Abschließend soll die Schnittstelle der beiden Müdigkeitserkennungen erklärt werden sowie der provisorische Algorithmus für den Prototypen zum endgültigen entscheiden, ob Müdigkeit vorliegt oder nicht.

3.1 Hardware

Für die Müdigkeitserkennung über Computer Vision wird ein Entwicklungsrechner und eine Kamera benötigt.

Der Entwicklungsrechner muss ein möglichst leistungsstarker, aber auch platzsparender Rechner sein, da dieser im Fahrsimulator möglichst wenig Platz in Anspruch nehmen sollte. Durch das Vorgängerprojekte [5][6] ist der Entwicklungsrechner für dieses Projekt allerdings schon vorgegeben. Der Entwicklungsrechner ist ein Nvidia Jetson TX2. Jedoch ist dieser Entwicklungsrechner sehr kostenintensiv. Daher soll der Jetson TX2 durch den Jetson Nano ersetzt werden. Dieser ist immer noch leistungsstark und kostet nur einen Bruchteil. Da der Nvidia Jetson Nano über eine eigene GPU (Grafikprozessor, engl. *Graphics Processing Unit*) verfügt, eignet sich dieser auch optimal für den Einsatz von Echtzeitbildverarbeitung. Ebenfalls ist der Entwicklungsrechner gut für den kombinierten Einsatz von EKG und Computer Vision geeignet, da die Müdigkeitserkennung über EKG hauptsächlich auf dem CPU (engl. *Central Processing Unit*) laufen wird, die Müdigkeitserkennung über Computer Vision auf der GPU.

Im Gegensatz zu Jetson TX2 verfügt der Jetson Nano nicht über eine interne Kamera. Daher wird eine externe Kamera benötigt, welche über den CSI-Ports (engl. *camera-serial-interface*) angeschlossen werden kann. Wichtig hierbei ist, dass nicht jede Kamera mit CSI verwendet werden kann. Diese unterscheiden sich beim Kamerasensor. Die meisten Kameras verwenden einen OV5647

Sensor, welche vom Jetson Nano nicht unterstützt werden. Stattdessen wird ein Modul mit einem Sony IMX219 Sensor benötigt. Das einzige Kameramodul, welches vom Jetson Nano unterstützt wird, ist somit das *Raspberry Pi Camera Module v2*.

Für die Erkennung des EKG-Signals wird die selbst entworfene Platine von [5] verwendet. Diese wird nach dem Platinenlayout nachgelötet. Anschließend kann diese auf einen Arduino Uno aufgesetzt werden (Abbildung 6).



Abbildung 6: Bestücktes EKG-Board aufgesetzt auf einem Arduino Uno [5].

3.2 Software

In diesem Abschnitt soll auf die Softwareentwicklung der Müdigkeitserkennung über Computer Vision und EKG eingegangen werden. Die verwendete Programmiersprache ist hierbei Python. Abschließend soll die Schnittstelle beider Müdigkeitserkennungen erklärt und auf den provisorischen Code zur endgültigen Müdigkeitserkennung eingegangen werden.

3.2.1 Computer Vision

In diesem Abschnitt sollen die verwendeten Bibliotheken und die Funktionsweise der Müdigkeitserkennung über Computer Vision erläutert werden.

3.2.1.1 Verwendete Bibliotheken

Open CV:

OpenCV ist eine kostenlose Bibliothek mit Algorithmen für die Bildverarbeitung und maschinelles Sehen. Sie ist für die Programmiersprachen C, C++, Python und Java geschrieben und steht als freie Software zur Verfügung. *OpenCV* wurde für die Recheneffizienz entwickelt und konzentriert sich auf die Bildverarbeitung in Echtzeit.

Dlib:

Dlib ist eine Bibliothek mit Algorithmen für maschinelles Lernen und Tools zum Erstellen komplexer Software in C++ und Python zur Lösung von Problemen der realen Welt. Es wird sowohl in der Industrie als auch in der Wissenschaft in einer Vielzahl von Bereichen eingesetzt, darunter Robotik, eingebettete Geräte, Mobiltelefone und große Hochleistungsrechnerumgebungen.

Aus *Dlib* wird die Funktion *Facial Landmark Detection* verwendet. Mit dieser Klasse ist es möglich, das Gesicht in definierte Punkte zu unterteilen (Abbildung 7).



Abbildung 7: Facial Landmark Detection⁵

⁵<https://www.pyimagesearch.com/2017/04/03/facial-landmarks-dlib-opencv-python/>, Zugriff: 13.03.20

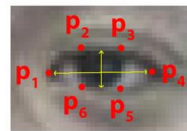
3.2.1.2 Funktionsweise und Schwellenwerte

Die Augen und der Mund werden mit Hilfe der Bibliotheken *OpenCV* und *Dlib* erkannt. *OpenCV* dient hierbei dem Einlesen der Kameradaten in Echtzeit, *Dlib* mit Hilfe von *Facial Landmark Detection* dem Aufteilen des Gesichts in spezifische Punkte. In Tabelle 2 sind die spezifischen Gesichtsregionen sowie die zugeordneten spezifischen Punkte mit *Facial Landmark Detection* zu entnehmen.

Tabelle 2: Gesichtsregionen und deren spezifische Punkte

Spezifische Punkte	Gesichtsregion
[1,17]	Kinnlinie
[18,22]	Rechte Augenbraue
[23,27]	Linke Augenbraue
[28,36]	Nase
[37,42]	Rechtes Auge
[43,48]	Linkes Auge
[49,68]	Mund

Mit Hilfe der spezifischen Punkte kann das Augen-Seitenverhältnis (engl. *eye aspect ratio* (EAR)) gebildet werden. Das Augen-Seitenverhältnis ist ein konstanter Wert, wenn das Auge geöffnet ist, fällt jedoch schnell gegen 0, wenn das Auge geschlossen ist [14]. In Abbildung 8 ist die Berechnung dieses Verhältnisses zu sehen.



$$EAR = \frac{\|p_2 - p_6\| + \|p_3 - p_5\|}{2\|p_1 - p_4\|}$$

Abbildung 8: Berechnung des Augen-Seitenverhältnisses⁶

⁶<https://hackaday.io/project/27552-blinktote-xt/log/68360-eye-blink-detection-algorithms>, Zugriff: 15.03.20

Zur Erkennung wie weit der Mund geöffnet ist, können die spezifischen Punkte 52 und 58 verwendet werden. Diese liegen direkt übereinander in der Mitte des Mundes (siehe Abbildung 7), daher muss hierfür kein spezielles Seitenverhältnis gebildet werden. Stattdessen muss nur der euklidische Abstand von Punkt 52 zu Punkt 58 berechnet werden.

Ziel der Müdigkeitserkennung über Computer Vision war das Erkennen von Müdigkeit der Stufe 3 nach der Wierwille-Skala (siehe Kapitel 2.1.1) und das Erkennen von Gähnen. Müdigkeit der Stufe 3 wird erkannt durch lange Lidschläge. Somit müssen für das Auge und den Mund jeweils zwei Schwellenwerte definiert werden. Ab welchem Wert gilt das Auge/ der Mund als geschlossen (augen EAR, Mund Abstand) und ab welcher Zeit (Augen Zeit, Mund Zeit) muss ein Müdigkeitsalarm ausgegeben werden. In [6] wurden diese Schwellenwerte bereits wie folgt definiert:

- Augen EAR: 0,3
- Augen Zeit: 900ms
- Mund Abstand: 25
- Mund Zeit: 3,5 Sekunden

In Abbildung 9 ist zu sehen, wie Müdigkeit aufgrund von Gähnen erkannt wird.

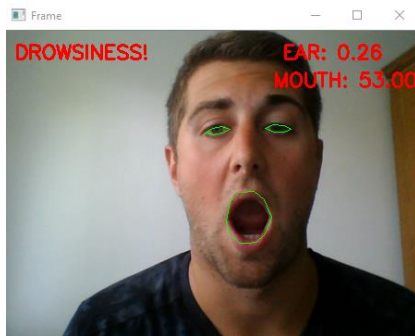


Abbildung 9: Müdigkeit durch Gähnen wird erkannt

Die rechts oben angezeigten Werte zeigten den berechneten Wert für das Augen-Seitenverhältnis sowie den euklidischen Abstand der Punkte des Mundes an. Die genaue Funktionsweise des Programms ist in [6] nachzulesen.

3.2.2 EKG-Erkennung

Für die EKG-Erkennung werden zwei verschiedene Software benötigt. Eine Software muss auf dem Arduino Uno laufen, die andere auf dem Jetson Nano.

Die Software auf dem Arduino liest über die analogen Eingänge das von der Platine aufbereitete EKG-Signal ein. Anschließend wird über die serielle Schnittstelle das EKG-Signal an den Jetson Nano gesendet.

Die Software des Jetson Nano empfängt das EKG-Signal über die serielle Schnittstelle. Bevor das EKG-Signal zur QRS-Komplex-Erkennung analysiert werden kann, muss dieses gefiltert werden. Dies ist notwendig, um diverse Störquellen des analog gemessenen EKG-Signals zu eliminieren. Beispielsweise bekommt man eine 50 Hz-Störung, sobald elektrische Leitungen in der Nähe sind. Das gefilterte Signal kann anschließend mittels adaptiver Schwellwert-Analyse und verschieden kleineren Filtern auf QRS-Komplexe untersucht werden, um daraufhin Herzrate bzw. Varianz derselben zu errechnen.

In Abbildung 10 sind die geplotteten Werte zu sehen. Links oben ist das ungefilterte Signal mit Störungen zu sehen. Links unten das gefilterte Signal. Auf der rechten Seite sind die Herzrate sowie die Herzfrequenzvariabilität zu sehen. Die genaue Funktionsweise der Filterung und Berechnung ist in [5] nachzulesen. Jedoch kommt es bei der dem verwendeten Programm aus [5] nach unbestimmter Zeit immer wieder zu unerklärlichen Abstürzen. Diese werden in Kapitel 4 genauer erläutert.

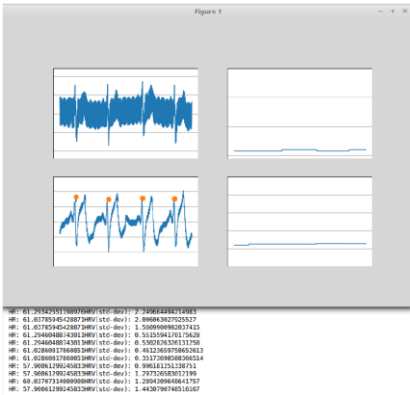


Abbildung 10: Plotten der gemessenen und erzeugten Signale [5]

3.2.3 Schnittstelle

Die Schnittstelle soll ein gleichzeitiges Starten der Müdigkeitserkennungen über EKG und Computervision ermöglichen. Für die Schnittstelle wird die Bibliothek *multiprocessing* verwendet.

Multiprocessing ist eine Bibliothek, welche das Starten von Prozessen mit Hilfe einer API unterstützt, die dem Threading-Modul ähnelt. Die Multiprocessing-Bibliothek bietet sowohl lokale als auch Remote-Parallelität, wodurch die globale Interpretersperre effektiv umgangen wird, indem Unterprozesse anstelle von Threads verwendet werden. Aus diesem Grund kann der Programmierer mit der Multiprocessing-Bibliothek mehrere Prozessorkerne auf dem Jetson Nano voll ausnutzen. Die Funktionen (Prozesse) werden somit parallel auf den zur Verfügung stehenden Prozessorkernen ausgeführt.

In Abbildung 11 ist der Zusammenhang zwischen den Python-Skripten zu sehen. Die Pfeile zeigen an, welches Skript Tasks aus den anderen Skripten startet.

Mithilfe von sogenannten *Queues* werden die zur Müdigkeitserkennung wichtigen Daten aus den Skripten von EKG und Computervision an die Schnittstelle übergeben. Dieses

wertet anschließend mit einem provisorischen Code die endgültige Müdigkeit aus.

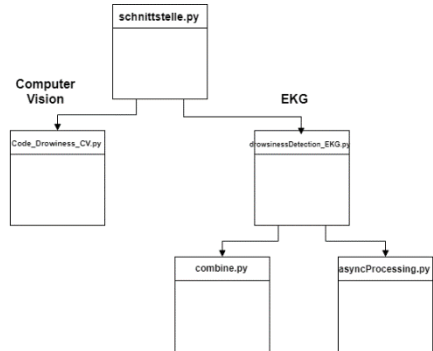


Abbildung 11: Pythonskripte und deren Zusammenhänge

Für den Code für die endgültige Müdigkeitserkennung werden Schwellwerte und einfache *if*-Anweisungen verwendet. Folgende Bedingungen für eine endgültige Müdigkeitserkennung gestellt (HR = Herzrate, HRV = Herzfrequenzvariabilität):

- $HR < 70$ und $HRV < 0,07$
- $HRV < 0,03$ und $HR < 75$
- Müdigkeit über CV erkannt und $HR < 75$
- Müdigkeit über CV erkannt und $HRV < 0,07$

Die Schwellenwerte für die Herzrate sind aufgrund der in Kapitel 2.2.2.1 wissenschaftlichen Literatur gewählt worden. Die Schwellenwerte für die Herzfrequenzvariabilität wurden auf Grundlage von [5] gewählt. Der Algorithmus wurde so einfach gewählt, da durch das Durchführen von Tests erste eigene Erkenntnisse zwischen Müdigkeit, Gesichtszügen und EKG gesammelt werden sollen. Auf diesen Erkenntnissen aufbauend soll der Algorithmus erweitert werden.

4 Experimente

Zum Gewinnen für Erkenntnisse für die Zusammenhänge zwischen Müdigkeit, Gesichtszügen und EKG sollen Experimente

durchgeführt werden. Hierbei soll mit vorläufig zehn Probanden die aktuelle Software getestet werden. Die Probanden sollen vor Beginn der Fahrt im Fahrsimulator einen Fragebogen ausfüllen, wie Sie Ihre Müdigkeit zu diesem Zeitpunkt einschätzen. Ebenfalls sollten diese vor dem Experiment keine wachmachenden Getränke wie Kaffee getrunken haben. Anschließend fahren die Probanden eine Stunde ohne Einflüsse von außen eine einfache Strecke. Dabei sollen Sie sich immer in Ihrer Fahrspur halten. Durch die Stunde monotone Fahrt soll der Proband müde werden. Abschließend soll von dem Probanden ein weiterer Fragebogen ausgefüllt werden, wie dieser während der Fahrt seine Müdigkeit eingeschätzt hat. Ebenfalls soll das Experiment mit einer Kamera aufgezeichnet werden, um im Anschluss das Experiment nochmals auswerten zu können. Für die Aufnahme des Videos kann nicht die angeschlossene Raspberry Pi Kamera verwendet werden, da dies erhebliche Performance-Einbußen zur Folge hätte.

Leider konnten die Experimente nicht durchgeführt werden. Anfangs stoppte das Programm von [5] nach maximal 10min. Dies konnte behoben werden, indem die in [5] verwendeten *Pipes* zur Datenübertragung durch *Queues*, welche wie bereits erwähnt auch im Programm der Schnittstelle verwendet werden, ersetzt wurden. Für einen Zeitraum von 5 bis 20min funktionierte die Software anschließend einwandfrei. Jedoch kommt es danach zu Fehlern in der QRS-Komplex-Erkennung. Es werden kaum noch bzw. gar keine QRS-Komplexe erkannt. Am geploteten Signal von Abbildung 10 ist jedoch ein einwandfreies EKG-Signal zu sehen. Eine Fehlersuche in der EKG-Software von [5] brachte ebenfalls keine Ergebnisse, da der Abbruch nach einer zufälligen Zeit stattfindet. Der Fehler muss bei der EKG-Software auf dem Jetson Nano sein, da das EKG-Signal über die serielle Schnittstelle beim Jetson Nano einwandfrei ankommt. Problematisch ist an der in [5] verwendeten QRS-Komplex

Erkennung, dass diese viel mit *Thresholds* arbeitet. Ändert sich die Stärke des Signals, kann es dazu führen, dass die Software den Komplex nicht mehr erkennt. Jedoch bringt eine Veränderung dieser *Thresholds* auch keinen Erfolg, da sich die Stärke des Signals dauerhaft ändert. Ebenfalls werden in dem Programm viele unkommentierte Buffer verwendet, deren Bedeutung meist unbekannt ist. Das macht eine Fehlersuche kaum möglich. Die Software scheint auf das in [5] gelötete Board abgestimmt zu sein. Dies macht eine Reproduktion der in [5] erzielten Ergebnisse kaum möglich. Daher empfiehlt es sich, eine neue EKG-Software auf Seiten des Jetson Nano zu schreiben, welche nicht auf ein Board abgestimmt ist, sondern allgemein aus einem EKG-Signal den QRS-Komplex und damit die Herzrate bestimmen kann.

5 Fazit und Ausblick

In dieser Arbeit wurde ein Prototyp zur Müdigkeitserkennung im Fahrsimulator entwickelt.

Zur Erkennung von Müdigkeit über Computer Vision wurden die Open Source Bibliotheken OpenCV und Dlib verwendet, für die Schnittstelle die Bibliothek multiprocessing. Die Müdigkeit wird mit Hilfe der Müdigkeitsbestimmung nach Wierwille und Ellsworth erkannt. Zur Erkennung von Müdigkeit über EKG wird eine von [5] selbst entworfene Platine sowie eine Software zum Auswerten verwendet. Die Müdigkeit wird mit Hilfe der Herzrate und der Herzfrequenzvariabilität erkannt. Bei der Herzfrequenzvariabilität wird das Verfahren der Standardabweichung (SDNN) verwendet. Für die Zukunft könnte dieses Verfahren durch die Spektralanalyse von LF und HF ersetzt werden. Dieses Verfahren ist im Bezug auf die Müdigkeitserkennung genauer und somit würden in Zukunft bessere Ergebnisse erzielt werden. Da die Software für die EKG-Erkennung neu programmiert werden sollte,

würde es sich anbieten, das Verfahren zu ändern.

Funktioniert die EKG-Software einwandfrei, können die Experimente durchgeführt werden, um erste Erkenntnisse zur Müdigkeitserkennung zu erhalten. Anschließend muss der provisorisch implementierte Algorithmus durch einen Algorithmus ersetzt werden, welcher auf den gewonnen Erkenntnissen aufbaut.

6 Literaturverzeichnis

- [1] K. T. Chui, K. F. Tsang, H. R. Chi, C. K. Wu, and B. W. Ling. 2015. Electrocardiogram based classifier for driver drowsiness detection. In 2015 IEEE 13th International Conference on Industrial Informatics (INDIN). Berlin Heidelberg New York, S. 600–603.
- [2] Jos'e Vicente, Pablo Laguna, Ariadna Bartra, and Raquel Bail'on. 2016. Drowsiness detection using heart rate variability. *Medical & Biological Engineering & Computing*.
- [3] S. Abtahi, B. Hariri, and S. Shirmohammadi. 2011. Driver drowsiness monitoring based on yawning detection. In 2011 IEEE International Instrumentation and Measurement Technology Conference. S. 1–4.
- [4] C Anitha, M Venkatesha, and B Suryanarayana Adiga. 2016. A Two Fold Expert System for Yawning Detection. *Procedia Computer Science* 92 (12 2016), S. 63–71.
- [5] M. Gromer, D. Salb. 2019. Müdigkeitserkennung am Steuer mittels EKG, Reutlingen.
- [6] S.Vennebusch. 2019. Müdigkeitserkennung über Computer Vision, Reutlingen.
- [7] W.W. Wierswille, L.A. Ellsworth. 1994. Evaluation of driver drowsiness by trained raters. *Accident Analysis and Prevention*, S. 571–581.
- [8] H. Kolrep, M. Rimini-Döring, A. Oehme, T. Jürgensohn, T. Altmüller. 2005. Wie sieht „müde“ aus? – Entwicklung und Validierung einer Skala zur Müdigkeitsbewertung von Kraftfahrern. In L. Urbas, & C. Steffens (Hrgs.), 6. Berliner Werkstatt Mensch-Maschine-Systeme, Zustandserkennung und Systemgestaltung, 13. – 15. Okt. 2005, Berlin, S. 65–70.
- [8] R. Provine. 2005. Yawning: the yawn is primal, unstoppable and contagious, revealing the evolutionary and neural basis of empathy and unconscious behavior. In: *American Scientist*. Band 93, Nr. 6 (November–Dezember), S. 532
- [9] A. Rosebrock. 2017. Facial landmarks with dlib, OpenCV, and Python
- [10] R. Klinge. 2011. Das Elektrokardiogramm - Leitfaden für Ausbildung und Praxis. Georg Thieme Verlag, Stuttgart.
- [11] A. Riztiane, D. H. Hareva, D. Stefani, S. Lukas. 2017. Driver Drowsiness Detection Using Visual Information on Android Device.
- [12] A. Y. Husodo. 2018. Multi-Parameter Measurement Tool of Heart Rate and Blood Pressure to Detect Indonesian Car Drivers Drowsiness.
- [13] S. Jung, H. Shin, W. Chung. 2012. Driver fatigue and drowsiness monitoring system with embedded electrocardiogram sensor on steering wheel.
- [14] R. Bhardwaj, P. Natrajan, V. Balasubramanian. 2018. Study to Determine the Effectiveness of Deep Learning Classifiers for ECG Based Driver Fatigue Classification.
- [15] A. Rosebrock. 2017. Facial landmarks with dlib, OpenCV, and Python.

User Interaction and Motion Dynamics Simulations for Microgravity – Development of a Concept and Prototype for Astronaut Training with Virtual Reality

Wafa M. Sadri
Reutlingen University
European Space Agency
Wafa.Sadri@iCloud.com

Abstract

Going forward with the requirements of missions to the Moon and further into deep space, the European Space Agency is investigating new methods of astronaut training that can help accelerate learning, increase availability and reduce complexity and cost in comparison to currently used methods. To achieve this, technologies such as virtual reality may be utilized. In this paper, an investigation into the benefits of using virtual reality as a means for extravehicular activity training in comparison to conventional training methods, such as neutral buoyancy pools is given. To help determine the requirements and current uses of virtual reality for extravehicular activity training first hand tests of currently available software as well as expert interviews are utilized. With this knowledge a concept is developed that may be used to further advance training methods in virtual reality.

Supervisor University: Prof. Dr. rer. nat.
Gabriela Tullius
Reutlingen University
gabriela.tullius@reutlingen-
university.de

Supervisor Organization: Dr. Andrea Casini
European Space Agency
andrea.casini@esa.int

Wissenschaftliche Vertiefungskonferenz
20. May 2020, Reutlingen University
Copyright 2020 Wafa Sadri

The resulting concept is used as a basis for development of a prototype to showcase user interactions and locomotion in microgravity simulations.

Keywords

Human Spaceflight; Virtual Reality; Astronaut Training; Learning; Educational Software

CR-Categories

Applied computing → Education → Interactive learning environments
Applied computing → Physical sciences and engineering → Aerospace
Human-centered computing → Interaction design → Interaction design theory, concepts and paradigms

1. Introduction

Construction for the International Space Station (ISS) began in 1998 and it was manned continuously from 2000 to today. With current plans dating the end-of-life of the ISS to the year 2030, space agencies like the National Aeronautics and Space Administration (NASA) and the European Space Agency (ESA) are starting to expand their human spaceflight programs beyond low Earth orbit (LEO). In preparation for these new missions, ESA is investigating methods of accelerating and streamlining astronaut training. One of these methods is the use of

Virtual Reality (VR) technologies. First, the context of the paper is described as well as the motivation behind the development of VR training simulations. Furthermore the goals of this work are outlined and the methodology used to achieve them. Finally, the current state of the art of microgravity motion and interaction simulation is examined and presented.

1.1. Motivation and Goals

For humans, space is a highly unusual environment, with varying gravity models ranging microgravity to reduced gravity levels on the Moon and Mars [1]. Furthermore psychological and physical stress induced by factors such as the dangers of spaceflight and living in close confinement with crew members of different nationalities pose challenges as well [1]. In order to train astronauts, many Earth based analog environments are utilized [1]. These analogs range from ISS module mockups to parabolic flights. Each of these environments convey varying levels of fidelity for different points of interest [1]. Currently, astronaut training for microgravity environment familiarization is carried out in two ways. The first method is the use of a neutral buoyancy pool. The focus of the neutral buoyancy training sessions, is to allow astronauts to get familiar with the locomotion in microgravity environments and learn the required tasks needed to carry out their extravehicular activity (EVA) [2]. In a neutral buoyancy pool, astronauts are usually equipped with a space suit or an analog thereof. They are then immersed into the pool, where mockups of ISS modules are featured. Using these mockups, a variety of tasks such as repairs, assembly and maintenance operations are trained [3]. However neutral buoyancy pools are limited in access, as they are often overbooked [2]. Furthermore Bolender et al. [2] describe that a training session usually requires the involvement of a Test Director, a Test Conductor, a Medical Doctor, a Safety Officer and an Audio/Video Operator. Additionally, there are at least one dive supervisor and six

supporting divers involved. This large amount of staff required to safely operate a training session increases the cost of a training session is significantly.

The second method for microgravity environment familiarization are parabolic flights. Here, an aircraft climbs steeply before entering a parabolic arc. In this configuration, the passengers and any payload will experience microgravity, until the aircraft climbs out of the dive at the end of the arc [4]. In this setting, weightlessness lasts between 20-30 seconds [4]. While the physical fidelity of this method is very close to effects experienced during a spaceflight, the relatively short duration of each parabola limits the time astronauts can spend learning locomotion and interaction with the simulated microgravity environment. Furthermore the cost of a parabolic flight starting upwards of \$5,000 per passenger is a factor to be considered as well¹.

In order to provide a more readily available and cost effective training method in comparison to both the neutral buoyancy pool and parabolic flights, VR technologies are proposed.

1.2. Methodology

In order to develop useful and accurate simulations for astronaut EVA training, first-hand experience from experts who will test the current state of the art of VR applications for microgravity simulations is utilized. Afterwards semi-structured expert interviews with closed and open questions will be held in order to capture test results as well as set requirements and evaluation goals for the interaction and locomotion concept and prototype to be developed in the course of this paper.

The objective of the expert interviews is to determine an outline of astronaut training as it is currently performed, what the goal of these training methods is, which tasks and interactions astronauts perform during EVAs and how they are limited by the environment and the EVA suit in performing these tasks and interactions. Furthermore, the interview

¹ gozerog.com/index.cfm?fuseaction=Reservations.welcome (Last access: 15. February 2020)

aims at discovering how and where VR training fits in the current training regime and what the expected benefits of VR astronaut training are. The experts consist of a range of astronaut instructors and engineers working at the European Astronaut Centre (EAC).

A total of 25 questions were defined by the author to determine an outline for a realistic training environment for development of a VR prototype. The questions were divided into two parts: The first part is focused on current EVA training and its goals as well as understanding specific task and interaction related aspects of EVAs.

The second part of the interview was aimed at the VR aspects of this work. To better understand the experts' opinions, as simple self-assessment of the interviewee's VR experience is requested. This is then followed by a series of questions related the most popular, currently available VR experiences that use microgravity locomotion in one form or another. The second half of questions aims to understand which aspects of EVA training may be best taught through VR training and why VR may be a suitable training tool. The final three questions determine the opinions of the interviewed experts on VR as a substitute or complementary method of training EVAs and where the added value of VR training lies.

The interviews were conducted in person with one interviewee and one interviewer respectively. The audio of the interviews was recorded temporarily and later transcribed in writing.

The ultimate scope of this work is indeed to investigate the locomotion and interaction techniques experienced and used by astronauts in real and simulated microgravity environments to define and outline a concept for a VR based motion dynamics simulation. The resulting concept may be used as a basis for development of a prototype to demonstrate the basic interaction and motion dynamics techniques. The development of the prototype will be carried out at the eXtended Reality Laboratory (XR Lab) laboratory at the EAC.

2. Basics

With the outline of this paper and methodology set, the basics for this work can be described. Human locomotion and interaction techniques used by astronauts in microgravity environments are described first. Second, the basic principles of human perception and learning, especially with regard to educational software and VR, are explored and analyzed. These are based on the results gained from the expert interviews.

2.1. *Human Locomotion and Interaction in Microgravity and Training Methods*

When humans move under the influence of a counterforce against gravity, such as the floor pushing up against a person, it is in our natural behavior to move by walking. However humans lack any motor skills for locomotion in microgravity, as true microgravity cannot be experienced on Earth. Even skydivers in free fall will experience drag which at terminal velocity acts as a counter-force large enough to stop the continuous acceleration towards the ground.

Therefor, in a microgravity environment, astronauts experience a highly unusual way of locomotion, mainly through the use of hand rails and foot restraints, which are found both inside and outside of the ISS [b]. Occasionally during an EVA, the astronauts are fixed to the end of the ISS Remote Manipulator System (SSRMS), in which case they are moved by a crew member from inside the ISS [b]. However the main methods of locomotion during EVAs consist of the use of hand rails for movement and tethers for safety [b]. For any self-induced movement, astronauts are typically tethered to the ISS with at least two tethers at any time. This is, because unlike the interior of the ISS, losing contact to the ISS would put astronauts in an extremely dangerous, potentially deadly situation [b]. In addition to these restrictions, astronauts need to keep their tools organized and in good working order while traversing on the exterior of the ISS [b].

While outside the ISS, astronauts are equipped with an EVA suit. Typically these

are the Russian Orlan or American Extravehicular Mobility Unit (EMU) space suits. The suits help protect the astronauts against the harsh space environment. They provide air pressure and oxygen supply high enough to support normal breathing as well as thermal control systems (TCS) to regulate body temperature. However, because of the protection they provide, they are bulky and difficult to use [b]. The movements of the space suit are heavily restricted and require considerable force to be performed [b]. Astronauts' abilities to move and interact with equipment and tools during EVAs are therefore heavily restricted. However the most profound effect on the interaction comes from the gloves [b], as these are the most used part of the EVA suit for interactions. One major constraint of the gloves is the significant reduction in haptic feedback to the astronauts [b]. This makes it difficult for astronauts to grab objects such as tools and handrails, as well as using tools. The latter is especially true due to the bulkiness of the gloves. In some cases it may prevent the hand from reaching into holes or underneath objects, where a gloveless hand would usually fit without issues [b]. However most repair and installation tasks that are performed during an EVA, including the tools used, are usually designed with these constraints in mind [a]. It is only when an unexpected failure occurs, that repair tasks not designed for EVAs have to be carried out [a]. The training for which usually takes place in the Neutral Buoyancy Facility (NBF) at the EAC or in the Neutral Buoyancy Laboratory (NBL) at Johnson Space Center (JSC) [a, b]. Here, astronauts wear special diving suits, similar to EVA suits used on the ISS. With these suits, limits in visibility and range of motion may be simulated [a]. The goal of these exercises is to familiarize astronauts with the EVA environment as well as the constraints they face while performing maintenance tasks outside the ISS [a, b]. Another method developed by NASA is the Active Response Gravity Offload System (ARGOS) [5]. ARGOS performs gravity offloading without having to immerse the astronaut into water. This allows for more accurate visual representations of the environment as well as

the tools used for the specific task. Development of EVA tools follows a process of prototyping and iteration, which includes constraints present for terrestrial testing environments such as neutral buoyancy pools, NASA Extreme Environment Mission Operations (NEEMO) and ARGOS [6]. The constraints range from waterproofing to weighting and the fact that achieving weightlessness for prolonged periods of time is unfeasible on Earth.

2.2. *Human perception and learning in the context of VR*

VR technology is mainly a visual medium. Therefore, using visual learning and memorization practices is advantageous for VR implementations of educational software. According to the definitions of teaching methods by Spalter et al. [7], VR technology would combine laboratory, visualization, simulation and lecture methods of teaching in one tool. While VR allows for immersion into the subject matter, this very immersion will often lead to significant work and cognitive overload in trainees [8]. A study conducted by Richards et al [9] concluded, that simply using VR as a means of training and teaching, does not inherently improve learning in itself. Such that simply presenting information on screen, whether this is on a PC monitor or in an immersive VR experience, does not implicitly improve learning [8, 9]. However both studies fail to make use of actual interactivity in the immersive medium. There are no methods for users to interact with objects within the scene, making the examined VR experiences simply passive video playback experiences, albeit in a 3D world with 6DOF.

Meyer et al. [10] shows that using immersive VR experiences for *pre-training* of trainees resulted in an overall significantly higher retention rate of gained knowledge, when compared to video lessons [10]. This suggests VR to be a useful tool for astronaut *pre-training and environment familiarization* and may be able to replace some currently used video and classroom lessons [b]. In fact, it is expected that 10-30% of all astronaut training may be conducted through VR technologies

in the near future [b].

One method of learning using visual methods is the memory palace technique, or Method of Loci (MOL), where a story or knowledge is built up through the use of an imaginary palace in which objects and symbols are placed in various rooms, each associated with a specific concept or idea [11]. With this method, early tests by Fassbender et al. [11] have shown that test subjects would retain information about 25% better when using the memory palace method as compared to a list of words. In a study with 78 participants conducted by Huttner et al. [12], users showed that in comparison to using a computer screen, users in VR would perform 5-7% better in retaining information. Furthermore the application of the MOL was significantly higher among VR users in that they were more inclined to make use of this learning method in comparison to the control group [12]. These results suggest that MOL and VR are well suited for each other.

A major benefit of VR technology is that it helps engage the trainee deep in the subject matter, provides new methods of presenting content and helps to preserve a consistent environment for learning for each trainee [13].

3. VR Applications for Microgravity Environments

Locomotion in microgravity environments has been explored in video games for a number of years now. For VR, there is a selection of games and experiences available, which allow the user to freely roam an environment in microgravity. In this section these games and experiences are examined, and their locomotive and interactive features with respect to EVAs are explored. Figure 1 shows the local coordinate system of an astronaut. Rotational motion around each axis is defined as follows: A rotation around the X as is pitching motion. Subsequently, a rotation around the Y axis is a roll maneuver and lastly a rotation around the Z is defined as yaw.

One of the earliest uses of VR technology for EVA training was the training used for the Hubble Space Telescope (HST) repair

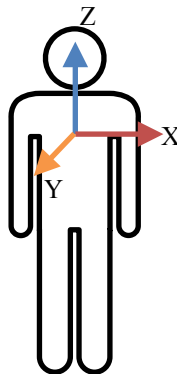


Figure 1: Local coordinate system of an astronaut [i]

mission in 1995 [14]. The objective of this training application was to teach flight team members EVA procedures and how to operate HST hardware for the repair mission [14]. To achieve this goal, 3D representations of both the Space Shuttle cargo bay and the HST were modeled [14]. While wearing a Head Mounted Display (HMD), the users would maneuver through the virtual world by using two joysticks, one for translating and the other for yaw, pitch and roll control, allowing for a full six degrees of freedom (DOF) [14]. Moreover, no locomotion using hand rails or tethering equipment was modeled. As the focus of this application lays within the realms of knowledge transfer for flight team members and not astronauts [14], this was also not necessary.

In early 2017 the VR experience Mission: ISS was released for the Oculus Rift. The experience allows the user to freely explore the interior of the ISS as well as carry out an EVA and spacecraft berthing. The user is given a basic set of controls and locomotion techniques. While inside and outside of the ISS, the user may use the thumb sticks on the controllers to move around, similar to a flying motion or a virtual jet pack. Additionally, the user can hold onto any surface, handrail and object with a collision box around it. By releasing the contact point while moving the hand, the user can generate a forward momentum in the opposite direction. Loose objects such as condiment containers, cargo bags and tools may also be interacted with.

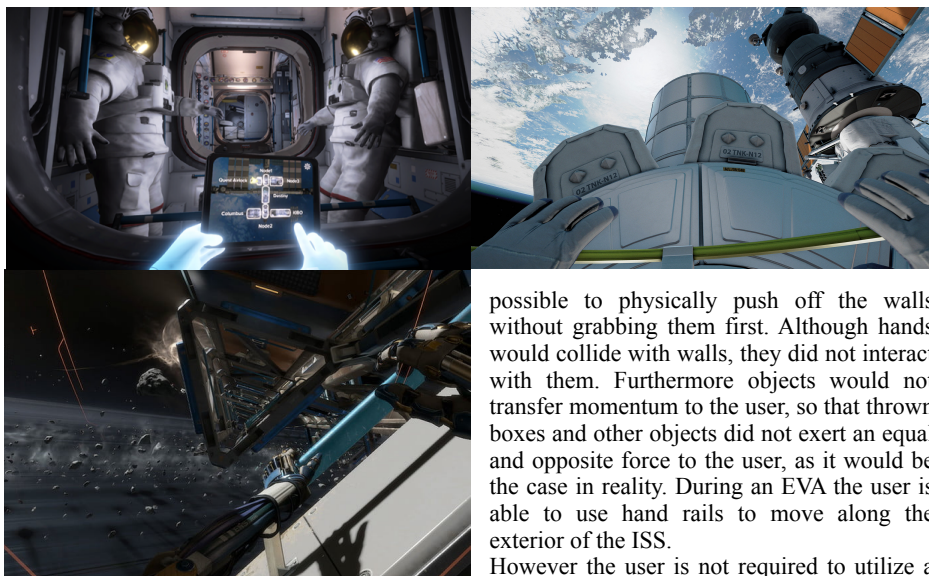


Figure 2: Mission: ISS (top left) [ii], BBC Home (top right) [iii], Lone Echo (bottom left) [iv]

The user can grab these objects and throw them. The exterior interaction follows a similar design. The user may hold onto hand rails as well as any collidable surface, while also being able to utilize a virtual jet pack to fly around the ISS. Locomotion is mainly based on the conservation of momentum. No roll control is provided to the users hand interaction as per the default setting. Much of the roll and pitch control is done by the second thumb stick control. However an optional setting is available that enables the user to use two hands for roll control. Yet no pitch control is available. The user may utilize pitch control by yawing 90 degrees to either direction and using a roll maneuver, then returning to the original yaw-orientation. The twin hand roll maneuver was perceived to be highly unpleasant during testing by experienced VR users and experts at the EAC's XR laboratory. The main complaint about the roll maneuver was related to its seemingly arbitrary point of rotation, which was perceived as unrealistic and confusing. At least one expert complained about headaches and nausea, directly related to the two handed roll maneuver. Furthermore it was not

possible to physically push off the walls without grabbing them first. Although hands would collide with walls, they did not interact with them. Furthermore objects would not transfer momentum to the user, so that thrown boxes and other objects did not exert an equal and opposite force to the user, as it would be the case in reality. During an EVA the user is able to use hand rails to move along the exterior of the ISS.

However the user is not required to utilize a tether and can freely float away and perform maneuvers with the Simplified Aid For EVA Rescue (SAFER) to fly around the ISS model. The SAFER system is a jet pack attachment to the EMU suit that allows astronauts to steer back to the ISS in case they become untethered and start drifting away.

A second EVA experience was provided by BBC Home - A VR Spacewalk, released in late 2017 for the Oculus Rift. Here, the user is set out to perform a virtual spacewalk or EVA on the ISS to assess damages. While the locomotion system is similar to the one used in Mission: ISS, it feels much slower and less responsive to user-induced motion. Unlike Mission: ISS, the user cannot hold on to walls and other static objects. The only objects that can be held onto are handrails. When moving, the inertia of the user is accounted for, making acceleration slower than the actual hand movement. The user is however not able to roll or pitch. The only possible user motions are lateral. The game will determine roll and pitch based on the players location. This means, roll and pitch are on rails. Whilst this is an acceptable solution for a guided VR experience for inexperienced users, it is unsuitable for astronaut training or familiarization, as it makes locomotion unrealistic and limited.

Lastly, the VR game Lone Echo was released for Oculus Rift in mid 2017. Here, the user is given an extensive locomotion mechanic. The user's virtual body, represented by an inverse-kinematic-influenced 3D model, interacts with the virtual environment in a realistic way where walls can be used to push off from, as well as hand rails to hold on to and change direction. However the user is not given the option to induce a roll movement. Therefore the directions up and down stay the same throughout the game. This helps significantly with orientation. Furthermore, it reduces risk for potential motion sickness.

4. VR Interaction and Locomotion Techniques for Microgravity Training

Based on the current state of the art and the experience gained from testing the applications previously mentioned as well as from expert interviews, a concept for VR interaction and locomotion for microgravity training is proposed. The concept presented here will be developed under the project name Virtual Reality Extravehicular Activity (VREVA).

4.1. Interaction

ESA has developed the Joint Investigation into Virtual Reality of Education (JIVE) as a complementary method of astronaut robotics training. Because JIVE [13] is already in the evaluation phase and is likely to be introduced as an option for astronaut training shortly, the interaction techniques used in VREVA are kept as close as possible to the ones used in JIVE. These are for example the layout and mapping of the buttons used for various functions. This is for familiarity reasons, so that astronauts do not need to learn the control mechanisms of the new application all over again, and can easily transition from one to the other and back. This means, the main way of interacting with the environment will be through motion controllers that the astronauts use with their hands. Feedback is provided to the astronaut via vibrotactile signals from the motion controllers. This way, the user can recognize interactable objects and surfaces and can react

accordingly. Providing this feedback to the user is a desirable feature that many of the tested games provide as well [d]. The feedback helps users recognize when an action can be performed and is the first step of the interaction between the virtual environment and the user. However it is not possible to keep all of the interaction designed for JIVE the same for VREVA. This is due to the constraint of microgravity locomotion. Where JIVE only utilizes one set of controls with a part of the controller using variable functions, this is not entirely possible in VREVA, because it utilizes two completely separate locomotion mechanisms. Therefore the general interaction using vibrotactile feedback for interactable surfaces and objects holds true, however the button mappings on the motion controllers differ when in microgravity-mode. Astronauts will ultimately learn to perform an EVA where they will use and interact with tethers, perform operations with tools and other equipment and control their space suit. These tasks include, but are not limited to, repair and maintenance work, installation of new systems and (passive) robotics operations.

4.2. Locomotion

While real microgravity locomotion allows astronauts to control all three rotational axes, roll, pitch and yaw, in VR this can lead to significant disorientation and motion sickness. As previously mentioned, roll maneuvers by the use of two hands were perceived as unpleasant by expert testers and the author, even leading to feelings of nausea and headaches. However, moving in this way is necessary and must be possible in the training simulator. Therefore it is desirable to allow users to adapt to this highly unusual form of locomotion by slowly introducing new techniques and maneuvers.

The main method of moving around outside of the ISS is through the use of hand rails. The astronaut can use their motion controllers to attach their virtual hands to these hand rails through the use of a grabbing action. By translating the hand in 3D space while grabbing, the astronaut is able to move themselves around the virtual space. If the astronaut lets go, the virtual body will

continue to move at the same velocity and in the same direction as it was in the moment the astronaut released the hand rail. This is due to the conservation of momentum. During translation up/down or left/right in the real world, astronauts would experience a change in pitch or yaw angle respectively. This is due to the fact that the center of mass is not aligned with the motion of the hand. However implicit counteracting of this pitch is performed by an astronaut to combat this. Because in VR trainees are not able to hold on to a physical handrail, this implicit counteracting does not occur. An expectation that the translation will not induce a change in pitch/yaw angle holds true.

There are multiple ways to perform more complex maneuvers such as pitch and roll. One of these is the use of two hands to induce roll. However as described previously, in the VR experience Mission: ISS this was perceived as highly discomfoting. To combat this, a dampening mechanism as well as a continuation of the translation mechanism may be used. For pitch and yaw, a two handed approach may also be useful, however it would also be possible to implement a yaw and pitching maneuver that increases with distance to the trainee's body. When the trainee is holding onto a handrail and their arm is close to their bodies, yaw and pitch maneuvers may be turned off completely, as to keep the expectations mentioned previously. As the trainee moves their arms further away from their bodies, pitch and yaw motions may be activated, increasing in effect for larger distances. Which of these approaches is most useful and least discomfoting, may only be evaluated after implementation and respective testing with trainees and experts.

4.3. Self-perception and embodiment

Humans perceive their own body at all times, mostly implicitly, but often actively. Most notable are hands and arms, as almost every task humans perform involve hands within the peripheral vision. Our bodies give us a lot of information regarding position in space and ability to perform actions [15]. Perceiving

your own body in VR can be done using an avatar, a virtual representation of a body. The amount of control over the avatar is directly linked to the feeling of presence within the simulation [15]. Furthermore presence is in turn directly linked to how well learned information is retained over extended periods of time [10]. For this reason it can be concluded that a well controlled avatar increases information retention rates for human locomotion training in microgravity environments. Therefore, an avatar is required for VREVA, that can be controlled by the trainee. This can be achieved in two ways. The first is using inverse kinematics (IK), where the location of the avatar's limbs, such as arms, elbows and legs, is calculated based on the position of the VR controllers. In the second approach, the limbs may be tracked with additional VR trackers. This method is more accurate in representing limb location however it involves more hardware and is therefore more costly and complex to use. A study by Schramm [15] finds that there is no significant difference in perceived embodiment from using IK versus tracking. While IK allows for joint limits to be set for a space suit, as this would be the case in reality, the same cannot be done with trackers in software alone. While trackers are more accurate, they would require an exoskeleton to act as a limiting mechanism for joint movements. While such an exoskeleton exists, it would need to be modified with trackers added to serve this function. Two arguments underline the use of exoskeleton. An exoskeleton provides accurate physical feedback to the trainee, which can provide an additional sense of realism. Secondly, its tracking markers can be fixed to the structure, allowing for more precise tracking of the limbs.

4.4. Learning

As outlined in chapter 2 an approach to VR education using MOL yields in good results for information retention. For this reason, VREVA will implement a similar concept to the ones mentioned previously. Additionally, the ESA developed JIVE training tool implements this principle as well [13]. JIVE presents the topics in different ways ranging

from self-study to goal-oriented approaches [13]. They are presented in an abstract form so that the knowledge gained by trainees is generic enough to be reused in future robotics projects beyond the ISS [13]. For VREVA a similar approach is chosen. A museum-like building structure is created, in order to fulfill the requirements of the MOL. In this, every room will represent a different subject matter, divided into stations to be completed by the trainee. Each lesson will introduce new concepts consecutively. Besides the locomotion and interaction aspects to be taught as mentioned in the previous sections, it is also important to teach the trainees about the environment [c]. This environment is the exterior of the ISS. While it is not important to model the ISS to the highest degree of detail in general, it is necessary to convey accurate representation of objects and surfaces astronauts will interact with, as they will later recall this information during their EVAs [b]. Furthermore, familiarization with the environment includes introduction to tools and techniques to be used during an EVA, as well as the EVA suit, which astronaut trainees will need to familiarize with.

4.5. Physiological aspects

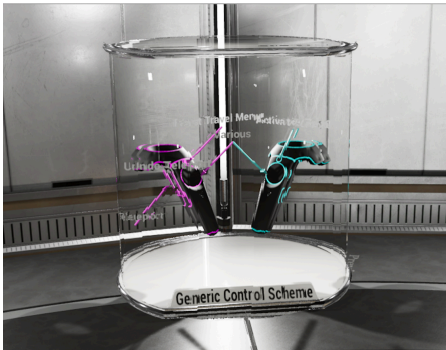
During VR training it is possible that trainees will experience discomfort with the equipment. Possible issues range from the heat of the HMD and possible receivers and batteries, to the overall weight resulting in neck strain [d]. Moreover the effects of VR sickness are well known today and range from simple discomfort and eye strain to extreme cases of headaches and even vomiting [16].

These factors have to be considered in the development of VREVA, as rooms and lessons have to be completable within an hour. It is possible that lessons in microgravity locomotion have to be considerably shorter than an hour, due to the fact that the unusual method of moving and the lack of inertial feedback can cause discomfort to a much higher degree than the teleportation locomotion present in JIVE [13]. To determine this, it is necessary to develop and evaluate a prototype that showcases the basic functionality.

5. Prototype Development

The concept presented in this work is used as a basis for development of a prototype for astronaut microgravity training, developed at ESA's XR Lab at the EAC. The prototype was developed using Unreal Engine 4.24 (UE4). In order to provide future projects with a platform, the development of the prototype followed a modular approach. The two main modules consist of generic items as well as project specific items. This guarantees that generic items such as the control and interaction scheme described in section 4 can be adopted by any future project. A development method like this is highly desirable among the interviewed experts, as it is a form of standardization that helps accelerate development processes in the future [b]. The hardware used for the development consisted of an HTC Vive Pro HMD and two controllers. Additionally, the HMD was equipped with a wireless setup.

As described previously the concept divides controls into the generic controls which are similar to the ones found in JIVE, and the microgravity mode controls. An unmapped button in the JIVE control scheme was used to allow switching between the two modes at any time. As described in the concept, the MOL will be used to teach the trainees. Three levels were built for this purpose. The main map, known as the hub, serves as the persistent map in UE4. The other two levels are a suit familiarization room and a microgravity testing room. Both are built to showcase the basic functionality of the prototype. The suit familiarization room implements the IK used for manipulating the suits joints. The general idea is that trainees are able to familiarize with the restrictions and joint limits of the space suit. A trainee may pick up the gloves and shoes of the suit to move them. Using IK the suit responds accordingly and the joints will bend as far as the limits will allow them. UE4 implements two types of IK solvers. The first being the FABRIK solver, developed by Aristidou et al. [17]. The second is an implementation of CCDIK as proposed by Wang et al. [18]. As the purpose of the space suit familiarization lesson is to allow trainees to experience joint



restrictions of a space suit, the CCDIK implementation was chosen, because it is the only available UE4 IK solver that implements joint limits.

The microgravity testing room consists of a small confinement for testing the locomotion and a basic mockup of the ISS module Columbus with handrails. In microgravity mode, the trainee is only able to grab handrails. Flat surfaces are cannot be grabbed. Yaw is provided to the user through the HMD, translation by using the hand controllers while grabbing onto a handrail. To grab a handrail, the user must squeeze the grip buttons of the HTC Vive controllers. Figure 3 shows the control scheme as it is presented to the user in the virtual environment. Pitch and roll have not yet been implemented for this prototype.

6. Discussion and Future Work

While neutral buoyancy training offers high fidelity in terms of locomotion, it lacks significantly in the visual aspect, as the training is carried out by mid to low fidelity mockups of the ISS modules [3]. In contrast, VR applications are capable of delivering high fidelity in the visual realm. However, in return they cannot deliver the same level of sophistication for locomotion simulation, since the user is confined to virtual objects. By using an exoskeleton, some physical feedback can be given to the trainee. For these reasons, a VR application may only be used as a complementary method of

training, alongside conventional training methods. However the availability and accessibility of VR is a significant factor. Therefore VR training may be able to replace some instructor lead classes and video based training. Another aspect to be considered when comparing neutral buoyancy and VR training is that EVA hardware is often designed and tested in neutral buoyancy pools [6]. In order to perform these tests, each tool must be physically built [6]. Using VR to test some design aspects of the tool may help accelerate its development, considering that the 3D Computer Aided Design (CAD) models created for the prototype manufacturing process may also be used in a VR environment.

The prototype developed in this paper shows the basic functionality of the concept defined. Continuing this work, an evaluation needs to be performed to find issues and improvements to the concept. Furthermore an extended feature set will have to be defined, including course work and lessons planned. Some of these may include generic familiarization with the EVA environment, the training of safety procedures such as tethering and rescue operations, generic maintenance tasks and tool usage [d]. After the review phase is completed, a full training simulator with all the functionality described in this concept.

7. Summary

Current astronaut training methods for microgravity involve a large number of instructors, supporting staff and specially equipped facilities. They need to be planned long in advance and are inflexible and costly. In order to mitigate these issues, complementary training methods such as VR training can be utilized. In this paper an overview of the requirements to develop a VR training system for microgravity is outlined by utilizing expert interviews. For these interviews 25 questions are defined that investigate different aspects of EVA training and VR applications. Furthermore, an analysis of human locomotion in microgravity and human learning in the context of VR educational applications is given. Using MOL

as a learning environment is determined to be a good fit to teach astronauts using VR. A test survey for current microgravity locomotion experiences is carried out. Three applications are then evaluated by focusing on their strengths and weaknesses.

The concept presented in this paper outlines and defines a method of implementing microgravity training simulators. It covers aspects from interactions and locomotion in microgravity, educational software in the context of VR and physiological aspects of trainees. A first prototype was developed using UE4, along with an HTC Vive Pro HMD and wireless setup. The prototype implements a control scheme similar to the one used for JIVE. This is done in order to provide control scheme familiarity across ESA's training simulations. User feedback on interactive surfaces is provided to the trainee via vibrotactile feedback from the HTC Vive controllers. Microgravity locomotion is performed through the use of hand rails. Translations are performed using the hand controllers. Yaw control is provided by the user. IK is used to provide a space suit joint limit familiarization demonstration. A full training simulator will be developed in the future and it will include all the findings of this work. Thus, a complete test will be done to assess the final result.

8. Acknowledgements

I would like to thank my EAC colleagues Anne Drepper, Rüdiger Seine, Stephane Ghiste and Lionel Ferra for supporting my work through participation in the interviews. Further I would like to thank Oliver Chard and Martial Costantini for their technical support and expertise which they kindly shared with me.

9. References

1. D.A. Vakoch et al.: On Orbit and Beyond - Psychological Perspectives on Humans Spaceflight; Second Edition; Springer-Verlag; Berlin, Germany; 2013; ISBN: 978-3-642-30583-2
2. H. Bolender, H. Stevenin, L. Bessone, A. Torres: Preparing for Space - EVA Training at the European Astronaut Centre; ESA Bulletin 128; Pages 33-40; ESA; Cologne, Germany; 2006; http://www.esa.int/esapub/bulletin/bulletin128/bul128e_bolender.pdf; Last access: 14. February 2020
3. J. Jairala, R. Durkin, Z. Ney, S. Parazynski: EVA Development and Verification Testing at NASA's Neutral Buoyancy Laboratory; NASA; Houston, TX; 2012; 20120006539
4. F. Karmali, M. Shelhamer: The dynamics of parabolic flight: Flight characteristics and passenger percepts; Acta Astronautica; Elsevier; 2006; doi: 10.1016/j.actaastro.2008.04.009
5. L.K. Dungan, P.S. Valle, D.R. Bankieris, A.P. Lieberman, L. Redden, C. Shy: Active Response Gravity Offload and Method; 2015; NASA; Washington D.C.; US Patent No.: US 9,194,977 B1
6. D. Budzyń, H. Stevenin, M. Maurer, F. Sauro, L. Bessone: Prototyping of Lunar Surface Geological Sampling Tools for Moon Spacewalk Simulations by ESA; European Space Agency; Italian Institute of Speleology, Bologna University; International Astronautical Congress; 2018; IAC-18,A5,1,12,x47143
7. A.M. Spalter, R.M. Simpson, M. Legrand, S. Taichi: Considering a full range of teaching techniques for use in interactive educational software: A practical guide and brainstorming session; Brown University, Dept. of Computer Science; IEEE Frontiers in Education Conference; Kansas City, MO; 2000
8. G. Makransky, T.S. Terkildsen, R.E. Mayer: Adding Immersive Virtual Reality to a Science Lab Simulation Causes More Presence But Less Learning; Dpt. of Psychology, University of Copenhagen; Psychological and Brain Sciences, University of California Santa Barbara; 2017
9. D. Richards, M. Taylor: A Comparison of Learning Gains when Using a 2D Simulation Tool Versus a 3D Virtual

- World: An Experiment to Find the Right Representation Involving the Marginal Value Theorem; Dpt. of Computing, Macquarie University; Journal of Computer & Education; Elsevier; 2015; <http://dx.doi.org/10.1016/j.compedu.2015.03.009>
10. O.A. Mayer, M.K. Omdahl, G. Makransky: Investigating the effect of pre-training when learning through immersive virtual reality and video: A media and methods experiment; University of Copenhagen, Dept. of Psychology; University of Southern Denmark, Dept. of Psychology; Elsevier; 2019; DOI: <https://doi.org/10.1016/j.compedu.2019.103603>
 11. E. Fassbener, W. Heiden: The Virtual Memory Palace; Journal of Computational Information Systems; Binary Information Press; 2006; 1553-9105
 12. J. Huttner, S. Robra-Bissantz: An Immersive Memory Palace: Supporting the Method of Loci with Virtual Reality; Institute of Technology, Brunswick, Germany; 23rd Americas Conference on Information Systems; Boston, MA; 2017
 13. S. Coene, L. Ferra, M. Costantini, A. Cowley, S. Ghiste: Development of a Virtual Reality Training Demonstrator at the European Astronaut Centre; ESA, European Astronaut Centre; International Astronautical Congress, Bremen, Germany; 2018; IAC-18-F1.2.3 IAC-18,B3,6-A5.3,8,x44462
 14. R.B. Loftin, P.J. Kenney: Training the Hubble Space Telescope Flight Team; Dept. of Natural Sciences, University of Houston-Downtown, TX; IEEE Computer Graphics and Applications; 1995; 0272-17-16/95
 15. R.C. Schramm: Vergleich von Inverse Kinematics und Motion Tracking im Zuge einer VR Anwendung; Bachelorthesis; Reutlingen University; 2019
 16. E.M. Kolasinski: Simulator Sickness in Virtual Environments; U.S. Army Research Institute; 1995; TR 1027
 17. A. Aristidou, J. Laseby: FABRIK: A Fast, Iterative Solver for the Inverse Kinematics Problem; Dpt. of Engineering, University of Cambridge; Graphical Models, Elsevier; DOI: 10.1016/j.gmod.2011.05.003
 18. L.T. Wang, C.C. Chen: A Combined Optimization Method for Solving the Inverse Kinematics Problem of Mechanical Manipulators; Dpt. of Mechanical Engineering and Technology, National Taiwan Institute of Technology; IEEE Transactions on Robotics and Automation; 1991; 1042-296X/91/0800-0489\$01.00
- ## 10. Interviews
- a. A. Drepper, interview by author, Cologne, 09. March 2020
 - b. R. Seine, interview by author, Cologne, 12. March 2020
 - c. S. Ghiste, interview by author, Cologne, 17. March 2020
 - d. L. Ferra, interview by author, Cologne, 02. April 2020
- ## 11. List of Figures
- i. Local Coordinate System of an Astronaut; Own representation
 - ii. Mission: ISS screenshot; <https://www.geekwire.com/2017/mission-iss-oculus-virtual-reality-space-station/>; Last access: 15. April 2020
 - iii. BBC Home - A Virtual Spacewalk screenshot; <https://www.godoculus.at/virtual-reality/home-gratis-vr-weltraumspaziergang-fuer-rift-vive/>; Last access: 15. April 2020
 - iv. Lone Echo screenshot; <https://www.vrfocus.com/2017/07/listen-to-lone-echos-inspiring-sci-fi-soundtrack/>; Last access: 15. April 2020
 - v. VREVA screenshot; Own representation

Vergleichende Analyse des YouTube-Auftritts von privat- und öffentlich-rechtlichen Sendegruppen

Alexander Kahabka

Hochschule Reutlingen

Alexander.Kahabka@Student.Reutlingen-University.de

Abstract

Lange wurde das Internet als Antagonismus zum Fernsehen gesehen. Es wurde dementsprechend zur Zuschauerrück- bzw. -gewinnung genutzt, was sich allerdings als ineffizient erwies. Inzwischen haben die einzelnen Sendegruppen das Internet jedoch als mediale Erweiterung erkannt und genutzt. Durch diese späte Akzeptanz zeigen sich starke Unterschiede im Umfang und der Vorgehensweise hinsichtlich der Nutzung des Internets als zusätzliches Medium. Am besten lässt sich dies in einem Vergleich in Bezug auf die wichtigste videotechnische Social Media Plattform YouTube darstellen.

In diesem Vergleich sollen die einzelnen Sendegruppen hinsichtlich ihrer wahrgenommenen Vorteile, Nachteile und Attraktivität bezogen auf das Nutzerverhalten und die Nutzermeinung bewertet werden. Die zielgruppenorientierte Optimierung des YouTube-Auftritts ist von außerordentlich hoher Bedeutung für die zukünftige Marktdurchdringung.

Schlüsselwörter

YouTube, Multimedia, Social media, Television, Public broadcasting, Commercial broadcasting

CR-Kategorien

• **Networks** → **Network types** → **Overlay and other logical network structures** → *Social media networks*; • **Information systems** → **Information systems applications** → **Multimedia information systems** → *Multimedia content creation*;

Einleitung

Anfang des zwanzigsten Jahrhunderts entstand das Fernsehen, Ende des zwanzigsten Jahrhunderts das Internet. Seitdem diskutiert die Medienindustrie über das Fortbestehen und die Entwicklung der beiden Technologien.

Seit der Entstehung der beiden Massenmedien hat sich einiges geändert. Die Befürchtung, dass das Internet das Fernsehen vollständig ablösen wird ist genauso ausgeblieben, wie vollständige Verdrängung des Radios durch das Fernsehen. Das Internet hat sich in den Jahrzehnten seit seiner Entstehung grundlegend geändert. Zu Beginn war es auf Grund der Komplexität nur Spezialisten möglich, Inhalte zu erstellen und zu bearbeiten. Es gab somit eine klare Trennung zwischen Produzenten und Konsumenten. Es wird hierbei vom Web 1.0 gesprochen. Im Laufe der Jahre wurde es immer einfacher, seinen Teil beizutragen und

Betreuer Hochschule: Prof. Terpinc
Hochschule Reutlingen
Boris.Terpinc@Reutlingen-
University.de

Wissenschaftliche Vertiefungskonferenz
20. Mai 2020, Hochschule Reutlingen
Copyright 2020 Alexander Kahabka

so entstanden immer mehr Content-Produzenten, genannt Web 2.0. Heutzutage ist jeder Nutzer ein Prosument, d.h. jedem Nutzer ist es möglich, Beiträge ins Internet zu stellen und diese abzurufen.¹

Aber auch das Fernsehen hat seit seinem Beginn eine deutliche Wandlung vollzogen. Besonders durch die Entstehung des Internets musste und muss in diesem Bereich weiterhin umgedacht werden. Die aktuelle Entwicklung zum neuen Fernsehen ist nicht zu übersehen. Früher war das Fernsehen ein passives Medium. Während die Zuschauer damals vor ihren Fernsehern saßen und zuschauten, geht dies heute nicht mehr so einfach. Das neue Fernsehen beschreibt die Möglichkeit, die von den Sendegruppen produzierten Inhalte plattformunabhängig wiederzugeben. Das Internet stellt hierbei keinen Antagonismus dar, sondern muss als eine mediale Erweiterung gesehen und genutzt werden. Hiervon machen die einzelnen Sendegruppen immer mehr Gebrauch. [1]

1.1 Motivation

Vor dem geschilderten Hintergrund stellt sich nun die Frage, ob und wie erwähnte Sendegruppen diese Möglichkeit nutzen. Um dies beantworten zu können, muss eine Social Media Plattform ausgewählt werden. Da es sich um Bewegtbilder handelt, bietet sich YouTube am ehesten an. Hierbei handelt es sich um eine der größten Social Media Plattformen und die größte Videostreaming Plattform². YouTube stellt somit das optimale Vergleichsmedium für Fernsehsender dar.

1.2 Problemstellung

Durch das schnelle Wachstum des Internets haben es die Sendegruppen verpasst, frühzeitig

ins Onlinegeschäft einzusteigen. Das trifft auch auf Social Media zu. Das Internet wurde hierbei lange Zeit als Konkurrenz gesehen. Es wurde überwiegend zur Rück- bzw. Gewinnung von Zuschauern genutzt, was jedoch nicht so effizient funktionierte, wie angenommen. Erst später wurde damit angefangen, das Internet als eigenständiges Medium zu akzeptieren und entsprechend mit Content zu versorgen. Der Umstieg auf das neue Fernsehen fand somit zeitlich verzögert statt. Auch wenn die zwei großen deutschen Sendegruppen gemeinsam zu spät in das Onlinegeschäft einstiegen, so unterscheiden sich doch die Herangehensweisen, besonders in Hinsicht auf YouTube.

1.3 Zielsetzung

Im Folgenden sollen nun diese Herangehensweisen der beiden Sendegruppen detailliert betrachtet werden. Zuerst soll aufgezeigt werden, wie sie vorgehen. Hierzu muss die Zielgruppe definiert werden. Außerdem muss die ursprüngliche Absicht, mit welcher die einzelnen Gruppen auf YouTube aktiv sind, beschrieben werden. Hieraus kann dann der Erfolg, anhand von Aufrufzahlen, Beliebtheit und anderen Metriken bestimmt werden. Ziel ist es hierbei, die beiden Gruppen bestmöglich miteinander zu vergleichen und für jede einzeln, sowie gemeinsam, Ergebnisse und Empfehlungen abzugeben.

2 Grundlagen

Dieses Kapitel beschreibt die beiden in Deutschland existierenden Sendegruppen. Diese werden jeweils einzeln beschrieben und analysiert. Anschließend werden sie miteinander verglichen und es werden besonders die Gemeinsamkeiten sowie Unterschiede aufge-

¹ Tim O'Reilly, *What Is Web 2.0*. [Online]. Available: <https://www.oreilly.com/pub/a/web2/archive/what-is-web-20.html> (accessed: Mar. 22 2020).

² We Are Social, Hootsuite, and DataReportal, *Digital 2020: Ranking der größten Social*

Networks und Messenger nach der Anzahl der Nutzer im Januar 2020. [Online]. Available: <https://datereportal.com/reports/digital-2020-global-digital-overview> (accessed: Mar. 19 2020).

zeigt. Darauf folgend wird mit der beschriebenen Einführung von Social Media die Online Plattform YouTube behandelt. Es wird die Entwicklung der letzten Jahre bis zur aktuellen Situation betrachtet.

2.1 Sendegruppen

Die Ursprünge der Fernstechnologie finden sich bereits im Jahr 1884, mit der sogenannten Nipkow-Scheibe. Hierbei handelte es sich um eine Spirallochscheibe, welche mechanisch Bilder in Hell-Dunkel-Signale aufteilen und diese ebenfalls wieder zusammenfügen konnte. 1928 wurde auf der 5. Großen Deutschen Funkausstellung in Berlin die erste richtige Fernsehübertragung durchgeführt.³

Bereits wenige Jahre später, am 22. März 1935, fand in derselben Stadt der erste regelmäßige Fernsehprogrammbetrieb, in von der Reichspost angelegten öffentlichen Fernsehstellen statt. In den folgenden Jahren entwickelte sich besonders in Deutschland, aber auch in anderen Ländern, das Fernsehen schnell weiter, nachdem sein Potential als Leitmedium klar wurde. Somit entstanden zuerst die ARD und später das ZDF. Erst 1984 kamen die beiden privaten Sendegruppen hinzu. [2]

2.1.1 Öffentlich-rechtlich

Nach dem Krieg wurde 1950 die ARD, also die Arbeitsgemeinschaft der öffentlich-rechtlichen Rundfunkanstalten der Bundesrepublik Deutschland, bestehend aus den sechs Landesrundfunkanstalten gegründet. Sie nahm 1952 den Fernsehbetrieb wieder auf. Nach mehreren

Namenswechseln heißt der gemeinschaftliche Sender Das Erste.⁴

In den folgenden Jahren konnten sich auf Grund der guten Wirtschaft immer mehr deutsche Bürger einen Fernseher leisten. Die auf Propaganda basierende Anfangsidee von Konrad Adenauer mit der Deutschland-Fernsehen GmbH wurde ihm untersagt. Hieraus entstand das ZDF, also das Zweite Deutsche Fernsehen, als eine von den bisherigen Anstalten unabhängige, zentrale und gemeinnützige Fernsehanstalt des öffentlichen Rechts. Es begann am 1. April 1963 mit der Übertragung. [2] Im darauffolgenden Jahr wurde das Fernsehen farbig.⁵

Da sich die beiden Sendegruppen somit in ihren Zielen sehr ähnlich sind, bilden sie zusammen die öffentlich-rechtlichen Sendegruppen. Ihre Ziele zur Qualitäts- und Leistungsorientierung lassen sich mit den ZDF-Leitlinien wiedergeben. Hier wird angegeben, dass es sich an den gemeinsamen Werten, Humanität, freiheitliche Demokratie, kulturelles Bewusstsein, sowie unabhängigem Journalismus orientiert. Die Verpflichtung liegt darin, den Zuschauern Verlässlichkeit und Respekt zu bieten. Diese Werte gelten ebenfalls für die programmliche Selbstverpflichtung, die Unternehmensposition und Dritte.⁶ Um dieses Angebot wirtschaftlich dauerhaft durchführen zu können, wurde kurz nach der Entstehung der ARD der Rundfunkbeitrag eingeführt. Dieser dient dazu, die öffentlich-rechtlichen Sendegruppen in allen Instanzen finanziell und wirtschaftlich abzusichern. Der

³ Norddeutscher Rundfunk, *Bewegte Bilder: Die Geschichte des Fernsehens*. [Online]. Available: <https://www.ndr.de/geschichte/Bewegte-Bilder-Die-Geschichte-des-Fernsehens,fernsehen240.html> (accessed: Apr. 12 2020).

⁴ Südwestrundfunk and ARD.de, *Das ist die ARD*. [Online]. Available: <http://www.ard.de/home/die-ard/organisation/>

Das ist die ARD/323058/index.html (accessed: Mar. 22 2020).

⁵ Matthias von Hellfeld, *ARD: 60 Jahre erfolgreich in Deutschland | DW | 05.06.2010*. [Online]. Available: <https://p.dw.com/p/Ngeh> (accessed: Mar. 22 2020).

⁶ Zweites Deutsches Fernsehen, *ZDF-Leitlinien*. [Online]. Available: <https://www.zdf.de/zdfunternehmen/zdf-leitlinien-100.html> (accessed: Mar. 22 2020).

Beitrag liegt heute bei 17,98 Euro und ist von jedem deutschen Haushalt monatlich zu entrichten. Die Einnahmen werden dann unter den einzelnen Rundfunksendern aufgeteilt.⁷

2.1.2 Privatrechtlich

In den darauffolgenden Jahren wurde die Fernsehtechnik immer weiterentwickelt und somit ursprüngliche Probleme, wie z.B. die Frequenzknappheit minimiert. Zusammen mit dem Wahlsieg der konservativen Parteien, welche oft Kritik am öffentlich-rechtlichen Fernsehen äußerten, und dem Ausbau der Kabelverbindungen durch die Bundespost führte dies zum Kabelpilotprojekt Ludwigshafen. So nahmen am 1. Januar 1984 SAT.1, damals PKS, und einen Tag später RTL, damals RTL plus, den Betrieb auf. In den folgenden Jahren wurde das zuerst nicht stark verbreitete Kabelnetz weiter ausgebaut, so dass Ende der Achtzigerjahre die Popularität der privaten Sendegruppen stark zugenommen hatte. Basierend auf Werbeeinnahmen konnte RTL bis 1991 seinen Brutto-Werbeumsatz auf über eine Milliarde und 1992 auf über zwei Milliarden DM beziffern. [3]

Inzwischen besitzen zwei Sendegruppen den überwiegenden Marktanteil. Diese sind die Mediengruppe RTL Deutschland GmbH, als Tochtergesellschaft der RTL Group, welche wiederum zu drei viertel Eigentum der Bertelsmann SE & Co. KGaA ist⁸, sowie die ProSiebenSat.1 TV Deutschland, als Holding-Gesellschaft der ProSiebenSat.1 Media SE⁹. Beide

besitzen abgesehen von dem aus dem Namen ersichtlichen Fernsehsendern ein großes, weiteres Angebot an Sendern.^{8,9}

2.1.3 Allgemeiner Vergleich

Im Zuge der Anfang der Neunzigerjahre steigenden Popularität der privatrechtlichen Sender und der unterschiedlichen Werbeschaltungen, öffentlich-rechtlich mit 20 Minuten pro Tag und privatrechtlich mit 288 Minuten pro Tag, geriet der Rundfunkbeitrag immer mehr in die Kritik. Das führte dazu, dass sich die öffentlich-rechtlichen Sendegruppen zurückwirkend wieder mehr auf Qualität- und Leistungsorientierung, mit Schwerpunkt Journalismus, fokussierten. [3]

2.2 YouTube

Der Begriff Social Media Network beschreibt die Möglichkeiten, sich über das Internet sozial miteinander zu vernetzen. Mit Hilfe dieser Vernetzung ist es möglich, untereinander zu kommunizieren und vom Nutzer erzeugte mediale Inhalte, sogenannten User-generated-Content, zu teilen. [4]

Das zweitgrößte soziale Netzwerk und die größte Videoplattform mit knapp 1,5 Milliarden Nutzern ist YouTube¹⁰. Der Grund für den Erfolg dieses Service ist die Mischung aus sozialem Netzwerk und Videoplattform. Jeder Nutzer kann Videos hochladen, welche dann wiederum von anderen Usern angeschaut, bewertet und kommentiert werden können. Da es jedem Benutzer möglich ist, Videos aller Art

⁷ Südwestrundfunk und ARD.de, *Entwicklung des Rundfunkbeitrags*. [Online]. Available: http://www.ard.de/home/die-ard/fakten/Entwicklung_des_Rundfunkbeitrags/1015754/index.html (accessed: Mar. 22 2020).

⁸ Mediengruppe RTL Deutschland GmbH, *Mediengruppe RTL Deutschland*. [Online]. Available: <https://www.mediengruppe-rtl.de/unternehmen/mediengruppe-rtl/> (accessed: Mar. 23 2020).

⁹ ProSiebenSat.1 Media SE, *Organisation und Konzernstruktur - ProSiebenSat.1 Media SE*

Geschäftsbericht 2017. [Online]. Available: <https://geschaeftsbericht2017.prosiebensat1.com/zusammengefasster-lagebericht/organisation-und-konzernstruktur.html> (accessed: Mar. 23 2020).

¹⁰ We Are Social, Hootsuite, and DataReportal, *Digital 2020: Ranking der größten Social Networks und Messenger nach der Anzahl der Nutzer im Januar 2020*. [Online]. Available: <https://datareportal.com/reports/digital-2020-global-digital-overview> (accessed: Mar. 19 2020).

hochzuladen, sofern diese nicht gegen die Richtlinien verstoßen, ist so gut wie jedes Genre auf YouTube zu finden. Auch die Machart der Videos unterscheidet sich stark. So finden sich von professionellen Serienproduktionen bis zu amateurhaften Spontanaufnahmen viele verschiedene Formate auf der Plattform. Interessant ist hierbei, dass YouTube bei der Erstellung eines Kontos nicht zwischen gewerblicher und privater Nutzung unterscheidet. Es besteht jedoch die Möglichkeit, ab einem bestimmten Bekanntheitsgrad, mit geschalteter Werbung in den eigenen Videos Geld zu verdienen. Neuerdings ist auch die Möglichkeit des Livestreamings nutzbar. Um bei der aktiven Suche und auch außerhalb höher gerankt zu werden, also z.B. in Vorschlägen oder Trends zu erscheinen, bieten sich u.a. eine hohe Aktivität, regelmäßiges Uploaden, guter Content und das Hinzufügen passender Metadaten an.¹¹

3 Vergleich

Nun sollen die beiden YouTube Auftritte miteinander verglichen werden. Hierbei ist der Umfang an Kanälen und deren Inhalt von Interesse. Außerdem soll gezeigt werden, was die Gruppen damit erreichen wollen und welche Unterschiede sich im Inhalt zeigen. Verglichen werden sollen auch die Zielgruppen und wie diese angesprochen werden. Daraus folgt dann die Beliebtheit der Gruppen. Mit den Ergebnissen soll anschließend eine Analyse stattfinden, welche die gravierenden Unterschiede hervorhebt und erörtert, ob und wie die gewünschten Ziele erreicht wurden.

¹¹ YouTube, *Suche und Auffindbarkeit auf YouTube - YouTube*. [Online]. Available: <https://creatoracademy.youtube.com/page/lesson/discovery?hl=de> (accessed: Mar. 19 2020).

3.1 Zielgruppen

In Abbildung 1 ist zu erkennen, dass ZDF bezogen auf den Marktanteil des Gesamtpublikums mit 12,7% und ARD mit 11,2% führend ist. Mit 8,5% folgen darauf RTL und mit 6,1% Sat.1. Setzt man den Fokus jedoch auf die Altersgruppe von 14 bis 49 Jahre, so führt hierbei RTL mit fast so viel Prozent wie ZDF im Gesamtpublikum, gefolgt von ProSieben, Sat.1 und VOX. Erst an vierter und fünfter Stelle folgen ARD und ZDF.

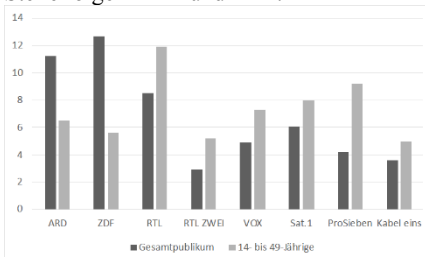


Abbildung 1: Marktanteile an Zuschauern nach Altersgruppen der acht größten Fernsehsender in Deutschland im Jahr 2019 nach¹²

Betrachtet man nun im Vergleich dazu, wie in Abbildung 2 zu sehen, die Altersverteilung der täglichen Nutzer von Internet, 54 Millionen, und Fernsehen, 48,9 Millionen, so fällt auf, dass der Trend hierbei genau gegenläufig ist. Je älter die Gruppe, umso höher ist der Fernsehnutzen und je jünger, umso höher ist der Internetnutzen.

Aus diesen Erkenntnissen lässt sich ableiten, dass die Herangehensweise der beiden Fernsehgruppen sehr unterschiedlich ist. Die privaten können dementsprechend ihr Fernsehprogramm direkt auf YouTube laden, da die Zielgruppen zwischen Fernsehen und

¹² David Grzeschik, *Der TV-Markt im Dezember: Das ZDF wird jünger, die meisten anderen verlieren*. [Online]. Available: <http://www.quotenmeter.de/n/114748/der-tv-markt-im-dezember-das-zdf-wird-juenger-die-meisten-andere-verlieren> (accessed: Mar. 25 2020).

Internet nicht stark voneinander abweichen. So haben die RTL-Gruppe, wie auch die ProSiebenSat.1Media jeweils direkt ihre eigenen YouTube Kanäle für die einzelnen Serien und Formate.

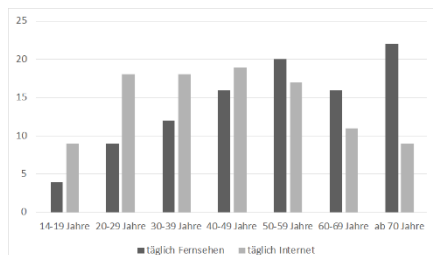


Abbildung 2: Prozentualer Anteil nach Altersgruppen an täglichen Nutzern des Fernsehens und des Internets nach [5]

Bei den öffentlich-rechtlichen ist die Vorgehensweise etwas differenzierter. Es würde sich hierbei nicht lohnen, das Fernsehprogramm, dessen Zuschauer ein Durchschnittsalter von über 60 Jahre aufweisen, direkt auf die Videoplattform zu laden. Zwar besitzen die öffentlich-rechtlichen ihre mit den Fernsehsendern verknüpften Kanäle, jedoch machen diese nicht den wichtigen und großen Teil aus. Aber auch hier gibt es erwähnenswerte Kanäle, wie z.B. der Kanal der Tagesschau mit mehr als einer halben Millionen Abonnenten¹³.

Den Beginn der öffentlich-rechtlichen machte 2009 das trimediale, also für Radio, Fernsehen und Internet geschaffene Jugendangebot *on3* des Bayerischen Rundfunks. Es wurde jedoch 2013 zu Puls umgeändert. Bereits *on3* besaß einen YouTube Kanal, welcher dann ebenfalls umbenannt wurde und heute *Puls Reportage*

heißt [6]. Neben *Puls Reportage*, siehe Abbildung 3, besitzt Puls nur noch einen weiteren Kanal, nämlich Puls Musik. Bis vor kurzem war die Anzahl an Kanälen noch höher, jedoch gehen immer mehr Kanäle, wie z.B. *Das schaffst du nie!* an Funk über¹⁴.

Funk ist das 2016 von den öffentlich-rechtlichen gemeinsam gegründete Content-Netzwerk für jüngere Zuschauer. Ziel ist es hierbei, den 14- bis 29-Jährigen ein öffentlich-rechtliches Internetangebot zur Verfügung zu stellen. Leitend ist bei dieser Gemeinschaftseinrichtung der SWR.¹⁵

3.2 Kanäle

Die Kanäle der einzelnen Sendegruppen lassen sich in Bezug auf ihren Ursprung in zwei Gruppen unterteilen. Die Kanäle können sender- oder programmbezogen sein. Senderbezogen bedeutet, dass einzelne Sender ihre eigenen YouTube-Kanäle haben. So besitzt das ZDF gerade einmal zwei, wohingegen die ARD je Sendeanstalt mindestens einen Kanal besitzt. Meist hat jeder Sender einer Sendegruppe einen eigenen Kanal. Diese haben jedoch, auf Grund der geringen Anzahl an Content, auch dementsprechend wenig Abonnenten. Anders ist dies bei den Programmbezogenen. Hier laden die jeweiligen Sendegruppen einzelne Folgen von Serien hoch. Da hier regelmäßiger Content hochgeladen wird, sind dementsprechend auch die Abonnentenzahlen größer. Bei der ARD ist hierbei, geleitet vom SWR und ausgestrahlt im Ersten, der Kanal *Verstehen Sie Spaß?* mit über eine Millionen Abonnenten erwähnenswert.¹⁶

¹³ ARD-aktuell, *tagesschau - YouTube*. [Online]. Available: <https://www.youtube.com/user/tagesschau> (accessed: Mar. 25 2020).

¹⁴ Südwestrundfunk, *Das schaffst du nie! - YouTube*. [Online]. Available: <https://www.youtube.com/channel/UC6Nh0xrtXJfNGucGtVrqjHw> (accessed: Apr. 12 2020).

¹⁵ Südwestrundfunk, *Transparenz*. [Online]. Available: <https://www.funk.net/transparenz> (accessed: Mar. 27 2020).

¹⁶ Südwestrundfunk, *Verstehen Sie Spaß?* [Online]. Available: <https://www.youtube.com/user/VSSpass> (accessed: Apr. 12 2020).

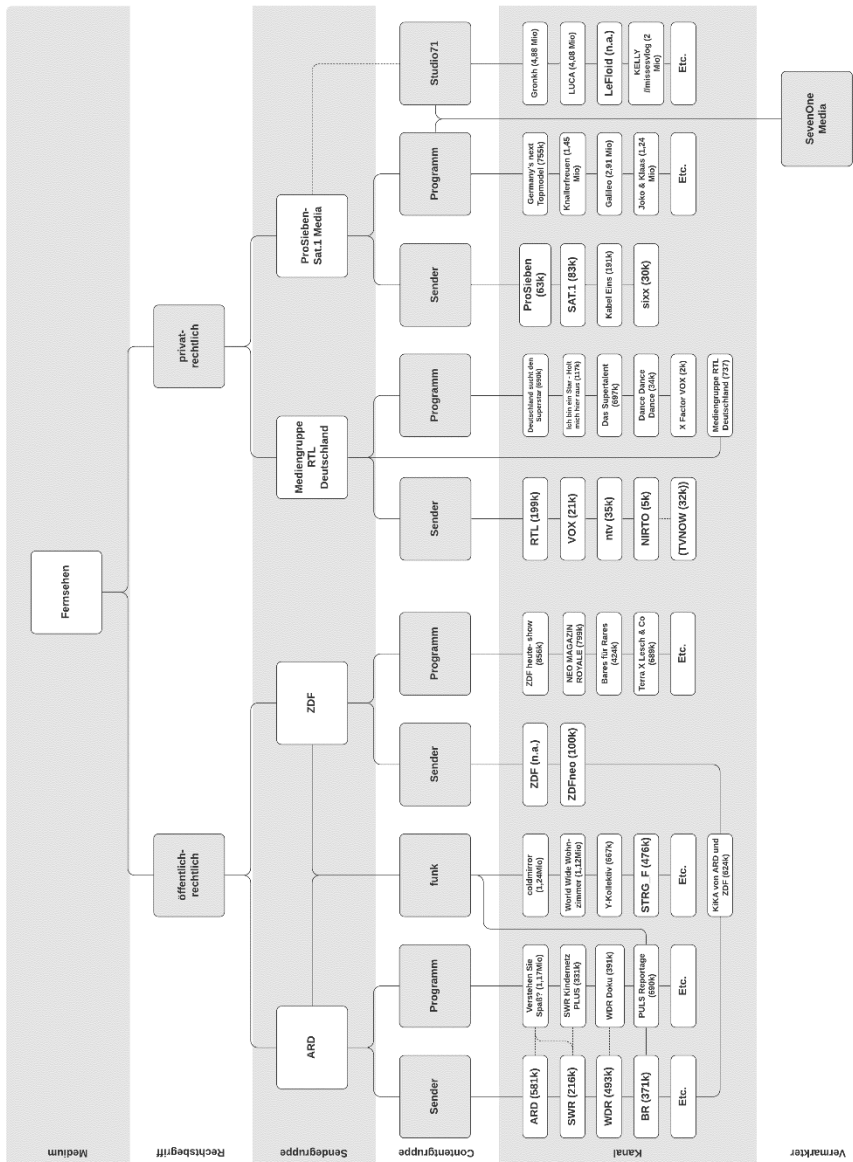


Abbildung 3: Struktur der Fernsehgruppen auf YouTube. Kanäle mit entsprechenden gerundeten Abonnentenzahlen¹⁷

Beim ZDF ist die *ZDF heute-show* mit fast neunhunderttausend Abonnenten führend.¹⁸ Das Contentnetzwerk Funk hingegen besitzt nach aktuellem Stand 72 Kanäle¹⁹. Diese sind im Gegensatz zu den anderen Kanälen der öffentlich-rechtlichen nicht alle nur durch bzw. für Funk entstanden. Somit ist zu unterscheiden zwischen selbst entstandenen und unter Vertrag stehenden Kanälen. So wird z.B. *STRG_F* vom Norddeutschen Rundfunk für Funk produziert,²⁰ wohingegen *coldmirror* bereits mehrere Jahre vor Funk auf YouTube aktiv war.²¹

Auch die privatrechtlichen gehen nach demselben Schema vor. Die Mediengruppe RTL Deutschland hat ebenfalls senderabhängige Kanäle. Auch für das Video-On-Demand-Angebot TVNOW existiert ein Kanal.²² Ebenfalls stehen für verschiedene Programmformate Kanäle bereit. Die Mediengruppe RTL Deutschland besitzt einen eigenen Kanal, welcher jedoch nur sehr wenig Abonnenten zählt.²³

ProSiebenSat.1 Media hat auch nach Contentgruppen sortierte Kanäle. Interessant ist hierbei der Bezug auf Studio71, welcher in 3.4 dementsprechend behandelt wird.

3.3 Inhalt

In Abbildung 3 lässt sich erkennen, dass sich besonders funk contenttechnisch abhebt. Während die restlichen Contentgruppen der Sendergruppen fast vollständig daraus bestehen, die medialen Fernsehinhalte unverändert sender- oder programmgebunden auf YouTube zu übertragen, bietet funk hierbei Formate, welche nur für die Plattform selbst entwickelt werden.²⁴

Die Kanäle von Funk richten sich bereits auf den ersten Blick an das von ihnen angegebene Zielpublikum. Die Formate werden überwiegend von sich ebenfalls im Zielgruppenalter befindlichen Creators produziert. Es dauerte einige Zeit, bis passende Formate entwickelt wurden. Die Anfangsformate setzten sich nicht durch. Viele der Creators, welche bereits vorher auf YouTube aktiv waren, wurden eingekauft und das Netzwerk wächst weiterhin. Somit sind auch alle Interessensbereich, von Comedy über Politik bis zu Aufklärung abgedeckt. Unter den Eigenproduktionen, welche speziell durch Funk aufgebaut wurden, besitzen die Kanäle für Reportagen die größte Beliebtheit. Mit diesen Kanälen kommt Funk

¹⁷ Abonnentenzahlen der gleichnamigen YouTube Kanäle von <https://www.youtube.com> entnommen (accessed: Mar. 27 2020).

¹⁸ Zweites Deutsches Fernsehen, *ZDF heute-show*. [Online]. Available: <https://www.youtube.com/channel/UCFqcNI0NAA21NS9W3ExCRg> (accessed: Apr. 12 2020).

¹⁹ Südwestrundfunk, *funk | Kanäle & Serien*. [Online]. Available: <https://www.funk.net/channel/> (accessed: Mar. 22 2020).

²⁰ Norddeutscher Rundfunk, *STRG_F*. [Online]. Available: <https://www.youtube.com/channel/UCfa7jJfYnn3P5LdJXsFkrjw> (accessed: Apr. 12 2020).

²¹ Kathrin Fricke, *coldmirror*. [Online]. Available: <https://www.youtube.com/user/coldmirror/about> (accessed: Apr. 12 2020).

²² Mediengruppe RTL Deutschland GmbH, *TVNOW*. [Online]. Available: <https://www.youtube.com/channel/UC58J5mgyNYw1IS947173qMw> (accessed: Apr. 12 2020).

²³ Mediengruppe RTL Deutschland GmbH, *Mediengruppe RTL Deutschland*. [Online]. Available: <https://www.youtube.com/channel/UC15HtB34MDtFdAzVcnOpQkw> (accessed: 12.4.2020).

²⁴ Südwestrundfunk, *funk*. [Online]. Available: <https://www.youtube.com/channel/UCOgPGtSnFR6GM-AkzCnxqMQ> (accessed: Apr. 12 2020).

dem Grundkonzept seiner Muttersendegruppen nach.²⁵

ARD und ZDF gehen hierbei sender- und programmbezogen vor. Hierbei dienen die Senderkanäle als Plattform für unterschiedliche Sendungen, sowie deren Aus- und Zusammenschnitte. Haben einzelne Sendungen genug Potential, so bekommen sie ebenfalls einen eigenen Kanal. So besitzt der allgemeine Kanal der ARD 652.000 Abonnenten²⁶ und der der *tagesschau* 584.000. Bei dieser Nachrichtensendung werden die einzelnen viertelstündigen Nachrichten ungeschnitten, sowie einzelne mehrminütige Beiträge geschnitten auf das Portal hochgeladen.²⁷

Abweichend von diesem Vorgehen ist *Verstehen Sie Spaß?*. Hierbei erscheinen nicht nur die im Ersten ausgestrahlten Folgen, sondern auch speziell für YouTube generierter Content auf dem Kanal. Dieser ist mit *online only* gekennzeichnet. In diesem Spezialcontent treten meist andere großen YouTuber auf.¹⁶ Durch das darauffolgende Crossmarketing, also das gegenseitige Bewerben von Kanälen, lässt sich die hohe Abonnentenzahl des Kanals erklären.

Ebenso wie ARD und ZDF gehen auch die Mediengruppe RTL Deutschland und ProSiebenSat.1 Media vor. Auch hier werden ganze Folgeanteile oder Zusammenschnitte gezeigt.²⁴ Somit spiegeln sich die Inhalte der einzelnen Sendegruppen auch in deren Content wider. Da die öffentlich-rechtlichen sich an bereits erwähnten gemeinsamen Werten orientieren, werden diese zum überwiegenden Teil auch in funk übernommen.

3.4 Beliebtheit

Die Beliebtheit der Plattformen lässt sich grundsätzlich an den Abonnentenzahlen und Videoaufrufen messen. Bei den öffentlich-rechtlichen besitzen die Senderkanäle jeweils mehrere hunderttausend Abonnenten. Auch die Programmkanäle befinden sich in diesem Bereich. Das ZDF besitzt jedoch mit der *ZDF heute-show* und *NEO MAGAZIN ROYALE* zwei Kanäle, welche auf die eine Millionen zugehen. Ähnlich wie bei *Verstehen Sie Spaß?* mit *online only*,¹⁶ besitzt auch das *NEO MAGAZIN ROYALE* Rubriken, welche als *webexclusive* nur über die ZDFmediathek oder YouTube abrufbar sind.²⁸

Die Abonnentenzahlen mancher Funk Kanäle sind deutlich höher. Dies liegt meist daran, dass die Kanäle bereits weit vor der Einführung von Funk entstanden sind. So ist z.B. *coldmirror* bereits seit 2006 auf YouTube aktiv.²⁰ Videos von Funk finden sich ebenfalls regelmäßig in den YouTube Trends wieder.²⁹ Die Mediengruppe RTL Deutschland lädt auf ihren Kanälen jeweils nur kurze Sequenzen hoch, ganze Folgen oder längere Zusammenschnitte sind hier weniger zu finden. Außerdem ist die Anzahl an Kanälen nicht sehr hoch und überschaubar.²⁴

Auch die senderbasierten Kanäle der ProSiebenSat.1 Media weisen keine starke Abonnentenzahl auf und folgen dem Beispiel der vorher genannten Gruppen. Bei einer genaueren Betrachtung von Abbildung 3 fällt jedoch die überdurchschnittliche Abonnentenzahl der Programmkanäle auf. Hier besitzen mehrere Kanäle über eine Millionen Abonnenten. Führend ist hierbei *Galileo* mit fast 3 Millionen

²⁵ siehe entsprechende YouTube-Kanäle in Abbildung 3

²⁶ Südwestrundfunk, *ARD*. [Online]. Available: <https://www.youtube.com/user/ARD> (accessed: Apr. 12 2020).

²⁷ Südwestrundfunk, *tagesschau*. [Online]. Available: <https://www.youtube.com/user/tagesschau> (accessed: Apr. 12 2020).

²⁸ Zweites Deutsches Fernsehen, *NEO MAGAZIN ROYALE - YouTube*. [Online]. Available: <https://www.youtube.com/channel/UCNNEMxGKV1LsKZRt4valbvw> (accessed: Apr. 12 2020).

²⁹ YouTube, *Trends*. [Online]. Available: <https://www.youtube.com/feed/trending> (accessed: Apr. 12 2020).

Abonnenten. Der Content ist zwar ebenfalls aus dem Fernsehen übernommen, jedoch besteht er meist aus einzelnen Beiträgen. Diese stellen für durchschnittliche YouTube Verhältnisse lange Sequenzen von über 10 Minuten dar. Auch die bereits 2015 eingestellte Sketchshow *Knallerfrauen* besitzt auf YouTube fast eineinhalb Millionen Abonnenten und ist regelmäßig in den Trends vertreten.³⁰

3.5 Ergebnisse

Die grundlegenden Vorgehensweisen der einzelnen Sendegruppen unterschieden sich nicht stark. Meist werden die für das Fernsehen produzierten Inhalte geschnitten, in kurzen oder langen Sequenzen auf das Internetportal hochgeladen. Contenttechnisch ist jedoch das Konzept der öffentlich-rechtlichen mit Funk sehr interessant.

Funks Kernprodukt ist die Videoproduktion [7]. Es etablierte sich auf dem Markt inzwischen so weit, dass fast drei viertel der Zielgruppe, also Nutzer zwischen 14- und 29 Jahren, die Formate bzw. das Netzwerk selbst kennen. Auch konnten Formate von Funk bereits verschiedene Preise gewinnen. Besonders im Fokus stehen hierbei die Informations- und Dokumentationskanäle. 2019 und 2018 waren unter den Nominierten für den YouTube Goldene Kamera Digital Award im Bereich *Best of Information* fast nur Funk Kanäle nominiert, von denen 2019 *STRG_F* den Preis gewann.³¹ Aufgrund dessen zeigt sich, dass das Vorgehen von Funk gut ist und seinen Zweck erfüllt. Die Strategie sollte jedoch angepasst werden, damit die Kanäle weiterhin bzw. noch stärker an Nutzern gewinnen.

Die Mediengruppe RTL Deutschland hingegen sollte definitiv mehr Wert darauflegen, Abonnenten zu gewinnen. Zwar besitzen die Kanäle auch genügend Aufrufe, jedoch kann

es auf Grund der geringen Abonnentenzahlen schnell dazu kommen, dass die Kanäle an Bekanntheit verlieren. Hier wären verstärktes Crossmarketing und die allgemeine verstärkte Anpassung der Inhalte für die Plattform empfehlenswert.

ProSiebenSat.1 Media ist von allen hier erwähnten Plattformen durch die 2013 gegründete Tochtergesellschaft Studio71 und deren Vermarkter SevenOne Media definitiv am besten aufgestellt. Nach der Gründung übernahm die Muttergesellschaft mit anderen internationalen Sendern das US-Netzwerk Collective Digital Studio und integrierte es daraufhin in Studio71. Die ProSiebenSat.1 Media behält somit drei viertel der Anteile. Dadurch wurde es zu einem der bekanntesten Netzwerke der Welt. Somit besitzt die SevenOne Media sehr viel Erfahrung in der Vermarktung von YouTube Kanälen. Dieses Potential wird auch auf die Programmkanäle der ProSiebenSat.1 Media angewendet. [8]

4 Fazit und Ausblick

Zusammenfassend lassen sich die Fernsehsender auf YouTube in drei Gruppen unterteilen: öffentlich-rechtlich, Mediengruppe RTL Deutschland und ProSiebenSat.1 Media.

Das Contentangebot und die Beliebtheit der Mediengruppe RTL schneidet im Verhältnis am schlechtesten ab. Die öffentlich-rechtlichen hingegen sind bereits umfangreicher und durchdachter aufgestellt. Ein großes Angebot an Kanälen und sogar ein eigenes Contentnetzwerk für junge Leute komplettieren das Gesamtangebot.

Führend ist ProSiebenSat.1 Media mit ihren programm-basierten Kanälen. Dass diese jedoch so beliebt sind, ist nicht verwunderlich, da das Marketing von dem Vermarkter einer der weltweit größten Webvideo-Netzwerke

³⁰ Studio71 GmbH, *Knallerfrauen*. [Online]. Available: <https://www.youtube.com/user/Knallerfrauen> (accessed: Apr. 12 2020).

³¹ Südwestrundfunk, *Drei Jahre funk: Drei Viertel der Zielgruppe kennt das Content-*

Netzwerk von ARD und ZDF. [Online]. Available: <https://www.swr.de/unternehmen/kommunikation/pressemeldungen/funk-drei-jahre-funk-100.html> (accessed: Mar. 29 2020).

übernommen wird.

Hieraus leitet sich ab, dass die Mediengruppe RTL Deutschland auf der Plattform zurückbleiben wird, sollte sie nicht in den nächsten Jahren ihre Strategie und das Marketing erneuern. Die öffentlich-rechtlichen sollten ebenfalls ihre Strategie regelmäßig anpassen und aktualisieren, damit sie den Anschluss nicht verpassen. Wahrscheinlich wird in Zukunft das Content-Netzwerk Funk weiter anwachsen und Zulauf von Creators, wie auch Zuschauern bekommen.

Favorit ist und wird auch in Zukunft ProSiebenSat.1 Media bleiben. Da diese Sendegruppe mit Studio71 und deren Vermarkter als Tochtergesellschaften unantastbar sind, haben sie dementsprechend die größte Reichweite und Beliebtheit.

5 Literaturverzeichnis

References

- [1] T. Deuerling, *Innovationsmanagement für neues Fernsehen*. Wiesbaden: Springer Fachmedien Wiesbaden, pp. 15-18, 2016, ISBN: 978-3-658-11668-2.
- [2] Carmen Thomas, “Kurze Geschichte des Fernsehens,” *Die Zeit*, 28 Dec., p. 42, 2006. https://www.zeit.de/2007/01/Kurze_Geschichte_des_Fernsehens/komplettansicht (accessed: Mar. 22 2020).
- [3] E. Karstens and J. Schütte, *Praxishandbuch Fernsehen*. Wiesbaden: Springer Fachmedien Wiesbaden, pp. 19-22, 2013, ISBN: 978-3-658-00113-1.
- [4] I. Sjurts, *Gabler Lexikon Medienwirtschaft*, 2nd ed. Wiesbaden: Gabler Verlag, 2011, ISBN: 978-3-8349-6487-8.
- [5] Beate Frees and Wolfgang Koch, “ARD/ZDF-Onlinestudie 2018: Zuwachs bei medialer Internetnutzung und Kommunikation,” *Media Perspektiven*, no. 9, pp. 398–413, 2018, ISSN: 0170-1754. [Online]. Available: http://www.ard-zdf-onlinestudie.de/files/2018/0918_Frees_Koch.pdf
- [6] K. Depta, “Rock- und Popmusik als Chance: Impulse für die praktische Theologie,” Dissertation, Institut für Katholische Theologie der Universität, Vechta, pp. 103-105, 2015, ISBN: 978-3-658-12188-4.
- [7] Sabine Feierabend, Pia Philippi, and Anna Pust-Petters, “funk – das Content-Netzwerk von ARD und ZDF,” *Media Perspektiven*, no. 1, pp. 10–15, 2018, ISSN: 0170-1754. [Online]. Available: https://www.ard-werbung.de/fileadmin/user_upload/media-perspektiven/pdf/2018/0118_Feierabend_Philippi_Pust-Peters.pdf
- [8] Anja Kruse-Anyaegbu, “ProSiebenSat.1, TF1 und Mediaset gründen europäische Digital-Allianz,” *Cable! Vision Europe*, 12 Jan., 2017. https://www.cablevision-europe.de/_rubric/detail.php?rubric=Markt%2BUnternehmen&nr=3625&PHPSESSID=10jt4il319p5p0pga3jthnrj72 (accessed: Mar. 29 2020).



Lust auf Zukunft?

**Karriere in der Medizintechnik –
in Tübingen und Rangendingen.**

→ Praxissemester, Abschlussarbeit oder Direkteinstieg
nach dem Studium?

**Beginnen Sie Ihre Zukunft bei uns.
www.erbe-med.com**

Erbe Elektromedizin GmbH Tübingen Deutschland +49 7071 755 0 erbe-med.com

20

Informatics Inside

INF



Hochschule Reutlingen
Reutlingen University