

**Menschliche Verhaltensanalyse zur Beurteilung
prozessbasierter Assistenzsysteme am Beispiel der
industriellen manuellen Montage**

Torben Töniges

Universität Bielefeld

Technische Fakultät

Angewandte Informatik

Research Institute for Cognition and Robotics (CoR-Lab)

NRW Fortschrittskolleg: Gestaltung von flexiblen Arbeitswelten

Menschliche Verhaltensanalyse zur Beurteilung prozessbasierter Assistenzsysteme am Beispiel der industriellen manuellen Montage

Torben Töniges

1. Gutachterin Prof. Dr.-Ing. Britta Wrede, Universität Bielefeld

2. Gutachter Prof. Dr.-Ing. Marc Hanheide, University of Lincoln

Prüfungsausschuss Prof. Dr.-Ing. Ulrich Rückert, Universität Bielefeld
Prof. Dr.-Ing. Britta Wrede, Universität Bielefeld
Prof. Dr.-Ing. Marc Hanheide, University of Lincoln
Dr.-Ing. Hendrik Buschmeier, Universität Bielefeld

Abdruck der genehmigten Dissertation zur Erlangung
des akademischen Grades Doktor-Ingenieur (Dr.-Ing.).
Der Technischen Fakultät der Universität Bielefeld
am 04. Dezember 2018 vorgelegt von Torben Töniges,
am 22. Februar 2019 verteidigt und genehmigt.

Torben Töniges

Universität Bielefeld

NRW Fortschrittskolleg: Gestaltung von flexiblen Arbeitswelten

Technische Fakultät

Universitätsstr. 25

33615 Bielefeld

E-Mail: ttoenige@techfak.uni-bielefeld.de

Gutachtende: Prof. Dr.-Ing. Britta Wrede und Prof. Dr.-Ing. Marc Hanheide

Prüfungsausschuss: Prof. Dr.-Ing. Ulrich Rückert, Prof. Dr.-Ing. Britta Wrede,

Prof. Dr.-Ing. Marc Hanheide und Dr.-Ing. Hendrik Buschmeier

Gedruckt auf alterungsbeständigem Papier nach DIN ISO 9706.

Eidesstattliche Erklärung

Hiermit erkläre ich, dass mir die geltende Promotionsordnung der technischen Fakultät bekannt ist. Ich habe die vorliegende Arbeit selbst angefertigt, keine Textabschnitte von Dritten oder eigener Prüfungsarbeiten ohne Kennzeichnung übernommen und alle von mir benutzten Hilfsmittel und Quellen in meiner Arbeit angegeben. Dritte haben weder unmittelbar noch mittelbar geldwerte Leistungen von mir für Vermittlungstätigkeiten oder für Arbeiten erhalten, die im Zusammenhang mit dem Inhalt der vorgelegten Dissertation stehen. Des Weiteren habe ich diese Dissertation noch nicht als Prüfungsarbeit für eine staatliche oder andere wissenschaftliche Prüfung eingereicht und ich habe keine gleiche, in wesentlichen Teilen ähnliche oder eine andere Abhandlung bei einer anderen Hochschule als Dissertation eingereicht. (siehe Rahmenpromotionsordnung §8 (1) g)

Bielefeld, 28. Februar 2019

Torben Töniges

Danksagungen

Diese Dissertation wäre ohne die Unterstützung von großartigen Menschen nicht möglich gewesen. An dieser Stelle möchte ich diesen Personen von ganzem Herzen danken.

Mama, Papa und Nikola – Ohne euch wäre ich nicht die Person, die ich heute bin. Ihr habt immer an mich geglaubt und mich immer unterstützt. Wir haben gemeinsam schwerste Zeiten gemeistert und sind zusammen stark geworden. Ich weiß, dass ihr viel zurückstecken musstet und dass auch ihr Unbeschreibliches durchgemacht habt. Ich werde euch nie zurückgeben können, was ihr für mich getan habt. Ohne euch würde ich diese Zeilen nicht schreiben können. Danke, dass ihr mir den nötigen Rückhalt gebt.

Nicola – Ohne dich hätte ich die abschließende Zeit meiner Promotion nicht geschafft. Du bist immer für mich da und weißt genau, wie du mich aufbauen kannst. Du hast mir den nötigen Optimismus gegeben weiterzumachen, hast mich motiviert und kritisiert, hast mich zum Lachen gebracht, hast für die nötige Ablenkung gesorgt und so manche emotionale Schiefelage gerade gerückt. Danke, dass du bei mir und für mich da bist.

Ich möchte meinen Freunden danken, auch wenn ich in der vergangenen Zeit gestresst war und wenig Zeit hatte, seid ihr immer für mich da gewesen.

Marc – Vielen Dank, dass du meine Arbeit als Gutachter bewertest.

Patrick, Sebastian, Nicola, Mama und Papa – Dankeschön, dass ihr mir ehrliches Feedback zu meiner Arbeit gegeben habt.

Britta und Franz – Vielen Dank, dass ihr mir die Chance gegeben habt, in eurer Arbeitsgruppe zu promovieren. Ich habe mich in der Arbeitsgruppe immer unterstützt und wohl gefühlt. Danke, dass ihr an mich geglaubt habt.

Raphael, Thorsten, Herr Dr. Buchalla und Herr Schareina – Dankeschön, für die tolle Zusammenarbeit und die Unterstützung. Ohne die Kooperation wäre diese Arbeit in dieser Form nicht möglich gewesen.

Ein ganz großes Dankeschön geht auch an die gesamte Arbeitsgruppe, das CoR-Lab, das Fortschrittskolleg und alle Personen, die mich während meiner Promotion begleitet haben und immer für einen Kaffee, Tee und Diskussionen zu haben waren.

Kurzfassung

In dieser Arbeit entwickle ich Verfahren, um während der Interaktion mit einem prozessbasierten System, einzelne Prozessschritte anhand des Nutzungsverhaltens zu bewerten und Störungen im Prozess zu erkennen. Darüber hinaus mache ich menschliche Verhaltensanalysen für weitere Entwicklungsprozesse von prozessbasierten Systemen verwendbar.

Ich integriere dazu verschiedene Arten der Prozessstörung in einem mithilfe von nutzerorientierten Entwicklungskonzepten erstellten Montageassistenzsystem, um natürliche menschliche Reaktionen auf Störungen auszulösen und somit die von mir entwickelten Verfahren der Erkennung dieser Schwierigkeiten zu evaluieren. Ich zeige, dass eine multimodale menschliche Verhaltensanalyse in der Lage ist, zusätzlich zu den durch die Versuchspersonen erkannten und verbalisierten schwerwiegenden Störungen weitere Verbesserungspotentiale im Prozess aufzuzeigen. Bei der Analyse menschlichen Verhaltens betrachte ich neben der direkten System-Interaktion und der Dauer einzelner Prozessschritte vor allem die Analyse des Gesichtes und der Kopfbewegung.

Ich entwickle sowohl überwachte als auch unüberwachte Erkennungsverfahren. Die Modellierung als ein binäres Entscheidungsproblem und die Evaluation dieses überwachten Verfahrens machen deutlich, dass eine Unterscheidung zwischen Prozessschritten mit schwerwiegenden Störungen und allen anderen Prozessschritten möglich ist. Die gelernten Modelle erzielen dabei zumeist nur befriedigende Leistungen. Bei der Verwendung von unüberwachten Methoden beschreibe ich ein neuartiges Verfahren, welches anhand eines gelernten Modells von „normalem menschlichen Verhalten“ während eines Prozessschrittes zuverlässig Prozessschritte mit abweichenden Verhaltensweisen identifizieren kann.

Durch die Anwendung der entwickelten Verfahren in einer digitalen Lernumgebung und somit einen deutlich anderen Bereich der Mensch-System Interaktion zeige ich schließlich die Generalisierbarkeit meiner entwickelten Methoden. Dabei sind die überwachten Verfahren in der Lage, menschliche Fehler vorherzusagen und die unüberwachten Verfahren können einzelne Aufgabenabschnitte in Hinblick auf ihre Schwierigkeit vergleichen.

Inhaltsverzeichnis

Abbildungsverzeichnis	xv
Tabellenverzeichnis	xvii
Abkürzungsverzeichnis	xix
1. Einleitung	1
1.1. Ziel dieser Arbeit	3
1.2. Aufbau dieser Arbeit	4
2. Hintergrund	5
2.1. Der Mensch in der industriellen Produktion der Zukunft	5
2.1.1. Automatisierungsszenario	7
2.1.2. Werkzeugszenario	8
2.1.3. Diskussion	9
2.2. Nutzerorientierte Entwicklung	10
2.2.1. Evaluationsmöglichkeiten	12
2.2.2. Diskussion	16
2.3. Gesichtsanalyse zur menschlichen Zustandserkennung	17
2.3.1. Formale Beschreibung der Mimik	18
2.3.2. Vorhandene praxisnahe Systeme	19
2.3.3. Diskussion	22
2.4. Zusammenfassung	23
3. Praxisnahe Anwendung der nutzerorientierten Entwicklungskonzepte	25
3.1. Projektverlauf	26
3.1.1. Modellierung des Montageprozesses	26
3.1.2. Entwicklung der Nutzeroberfläche	32
3.1.3. Aufbau des finalen Assistenzsystems	35
3.2. Evaluation	36

3.3. Zusammenfassung	37
4. Merkmalsdefinition zur Verhaltensanalyse in Montageumgebungen	39
4.1. Merkmalsgruppen auf Basis von Kopfausdrücken	40
4.1.1. Kopfausdruckerkennung	40
4.1.2. Aufmerksamkeitserkennung	51
4.1.3. Merkmalsgruppendefinition (zeitlich normiert)	53
4.2. Merkmalsgruppen auf Basis von System-Events	56
4.3. Merkmalsgruppe zur Berücksichtigung zeitlicher Variationen	58
4.4. Zusammenfassung	58
5. Vom praktischen System zur Evaluationsplattform	61
5.1. Anpassungen des Assistenzsystems	61
5.1.1. Integration eines Kamerasystems	62
5.1.2. Anpassung an unerfahrene Versuchspersonen	63
5.1.3. Manuelle Einführung von Störungen im Prozessablauf	63
5.2. Beschreibung der Datenerhebung	70
5.3. Auswertung der Fragebögen	73
5.4. Zusammenfassung	75
6. Erkennung von Prozessstörungen durch überwachte Verfahren	77
6.1. Systembeschreibung	78
6.1.1. Optimierung	78
6.1.2. Validierung	83
6.2. Auswertung	87
6.2.1. Action Unit Auftreten vs. Action Unit Intensität	89
6.2.2. Vergleich der Klassifikationsmodelle	90
6.2.3. Kombination verschiedener Merkmalsgruppen	91
6.2.4. Qualitative Fehleranalyse	96
6.3. Zusammenfassung und Diskussion	101
7. Erkennung von Prozessstörungen durch unüberwachte Verfahren	105
7.1. Systemaufbau	106
7.2. Auswertung	107
7.2.1. Verwendung aller Merkmalsgruppen	108
7.2.2. Verwendung der Gesichtsausdrücke	112
7.2.3. Verwendung der System-Events	114

7.2.4. Verwendung der Aufmerksamkeit	116
7.2.5. Verwendung der Dauer	117
7.3. Zusammenfassung und Diskussion	118
8. Methodentransfer	121
8.1. Studienbeschreibung	122
8.2. Anwendung der überwachten Verfahren	124
8.3. Anwendung der unüberwachten Verfahren	126
8.4. Zusammenfassung und Diskussion	130
9. Fazit	133
9.1. Zusammenfassung	134
9.2. Ausblick	136
A. Zusätzliches Material der Datenerhebung in der Montage	139
A.1. Studienmaterial	139
A.1.1. Stellungnahme der Ethik-Kommission	139
A.1.2. Einverständniserklärung	140
A.1.3. System Usability Score (SUS) Fragebogen	141
A.1.4. Raw Task Load Index (RTLX) Fragebogen	142
A.1.5. Persönlichkeitsfragebogen	143
A.2. ROC-Kurven unterschiedlicher Merkmalskombinationen	146
B. Zusätzliches Material der Studie zum Methodentransfer	155
B.1. Studienmaterial	155
B.1.1. Stellungnahme der Ethik-Kommission	155
B.1.2. Studienbeschreibung	156
B.1.3. Einverständniserklärung	157
B.1.4. Persönlichkeitsfragebogen	159
C. Literaturverzeichnis	161

Abbildungsverzeichnis

2.1. Nutzerorientierter Entwicklungsprozess	10
3.1. Aufgabentypen nach BPMN 2.0 Standard	28
3.2. Beispielhafter sequentieller Prozess	28
3.3. Schwimmbadmodellierung nach BPMN 2.0 Standard	29
3.4. Subprozessmodellierung nach BPMN 2.0 Standard	30
3.5. Prozessablauf des modellierten Anwendungsszenarios	31
3.6. Nutzeroberfläche des ersten Prototypen	33
3.7. Nutzeroberfläche des finalen Systems	33
3.8. Bildvergrößerung in der Nutzeroberfläche	34
3.9. Beschreibungsarten eines Montageschrittes	34
3.10. Zugriff auf die Dokumentation in der Nutzeroberfläche	35
3.11. Finale Version des entwickelten Arbeitsplatzes	36
4.1. Kopfausdruckererkennung und Merkmalsgruppenextraktion	41
4.2. Ergebnis der Vorverarbeitung	42
4.3. Vorverarbeitung der Kopfausdruckererkennung	43
4.4. Zwischenschritte der Kopfausdruckererkennung	44
4.5. Beispiele gezeigter Kopfausdrücke bei der Systeminteraktion	48
4.6. Auftretende Erkennungsfehler der Vorverarbeitung	49
4.7. Definierte Aufmerksamkeitsbereiche	52
4.8. Merkmalsextraktion auf Basis von System-Events	56
4.9. Verwendete Merkmale, Merkmalsgruppen und deren Ursprung	59
5.1. Kamerapositionen im Montageassistenzsystem	62
5.2. Prozessstörung im 8. Schritt des Montageprozesses	65
5.3. Prozessstörung im 35. Schritt des Montageprozesses	66
5.4. Prozessstörung im 23. Schritt des Montageprozesses	67
5.5. Schritt 58, in dem der Fehler aus Schritt 23 realisiert wird	68
5.6. Prozessstörung im 84. Schritt des Montageprozesses	68

5.7.	Prozessstörung im 11. Schritt des Montageprozesses	70
5.8.	Prozessstörung im 60. Schritt des Montageprozesses	71
5.9.	Prozessstörung im 62. Schritt des Montageprozesses	72
6.1.	Überwachtes Klassifikationssystem	79
6.2.	Ergebnisse bei Variation der Klassifikationsketten und Merkmalsgruppen . .	87
6.3.	Vergleich Action Unit Auftreten und Action Unit Intensität	89
6.4.	Vergleich der Klassifikationsmodelle	90
6.5.	Vergleich der Merkmalsgruppenkombinationen	92
6.6.	Vergleich und Signifikanzen von Merkmalsgruppenkombinationen	95
6.7.	Häufigkeit einer falschen Klassifikation je Prozessschritt	97
6.8.	20. Schritt des Montageprozesses	98
6.9.	61. Schritt des Montageprozesses	99
6.10.	18. Schritt des Montageprozesses	100
6.11.	45. Schritt des Montageprozesses	100
7.1.	Verfahren zum Erlernen eines Modells „normalen Verhaltens“	107
7.2.	Distanzberechnung mit dem gelernten Modell	108
7.3.	Distanzen-Heatmap bei Verwendung aller Merkmalsgruppen	108
7.4.	31. Schritt des Montageprozesses	109
7.5.	52. Schritt des Montageprozesses	110
7.6.	55. Schritt des Montageprozesses	111
7.7.	Distanzen-Heatmap bei Verwendung der Action Units	112
7.8.	9. Schritt des Montageprozesses	113
7.9.	Distanzen-Heatmap bei Verwendung der System-Events	114
7.10.	14. Schritt des Montageprozesses	114
7.11.	79. Schritt des Montageprozesses	115
7.12.	Distanzen-Heatmap bei Verwendung der Aufmerksamkeit	116
7.13.	Distanzen-Heatmap bei Verwendung der Dauer	118
8.1.	Vergleich zwischen Level 1 und Level 8 des Spiels „Give Up 2“	122
8.2.	ROC-Diagramm der besten Klassifikationskette	125
8.3.	Distanzen-Heatmap bei Verwendung der Action Units	126
8.4.	Vergleich zwischen Level 1 und Level 2 des Spiels „Give Up 2“	127
8.5.	Vergleich zwischen Level 6 und Level 7 des Spiels „Give Up 2“	128
8.6.	Distanzen-Heatmap bei Verwendung der Levelversuche	129

Tabellenverzeichnis

3.1. Prozessdaten des Anwendungsszenarios	32
4.1. Automatisch erkennbare Action Units	47
5.1. Teilnehmendendaten der Datenerhebung	73
5.2. System Usability Score (SUS) des entwickelten Assistenzsystems	74
5.3. Raw Task Load Index (RTLX) Auswertung der Datenerhebung	74
6.1. Übersicht aller ermittelten durchschnittlichen AUROC-Werte	88
6.2. Rangordnung der Klassifikationsmodelle	91
6.3. Rangordnung der Merkmalsgruppenkombinationen	93
6.4. Signifikanzen aller Merkmalsgruppenkombinationen	94

Abkürzungsverzeichnis

AAZ	Arbeitsablauf-Zeitanalyse
AU	Action Unit
AUROC	Area under Receiver-Operating-Characteristic Curve
BPMN 2.0	Business Process Model and Notation 2.0
CLM	Constrained Local Model
CNF	Conditional Neural Field
CNN	Convolutional Neural Network
FACS	Facial Action Coding System
FAR-Test	Friedman Aligned-Rank Test
FP-Analyse	Finner Post-hoc Analyse
FPR	Falsch-Positiv-Rate
HANNING	HANNING ELEKTRO-WERKE GmbH & Co. KG
HOG	Histogram of Oriented Gradients
IMEMO	Interaktive Mitarbeiterassistentz für die Elektromotorenmontage
it's OWL	Technologie-Netzwerk Intelligente Technische Systeme Ost-WestfalenLippe
LGOCV	Leave One Group Out Kreuzvalidierung
MCD	Minimum Covariance Determinant
MMOD	Max-Margin Objekt Detektion
MTM	Methods-Time Measurement

NASA-TLX	NASA-Task Load Index
NPS	Net Promoter Score
PCA	Principal Component Analysis (Hauptkomponentenanalyse)
PDM	Point Distribution Model
PS	Procrustes Superimposition
REFA	Verband für Arbeitsgestaltung, Betriebsorganisation und Unternehmensentwicklung
RLMS	Regularised Landmark Mean Shift
ROC	Receiver-Operating-Characteristic
RPR	Richtig-Positiv-Rate
RTLX	Raw Task Load Index
SUS	System Usability Score
SVM	Support Vector Machine
SVR	Support Vector Regression
SWAT	Subjective Workload Assessment Technique
TMU	Time Measurement Unit
TPOT	Tree-based Pipeline Optimization Tool
UEQ	User Experience Questionnaire
UMUE	Usability Metric for User Experience
WSR-Test	Wilcoxon Signed-Rank Test

Einleitung

„ *Innovation distinguishes between a leader and a follower.*

– Steve Jobs

Nachdem in der digitalen Transformation der Industrie lange die Automatisierung und Vernetzung von Produktionsanlagen im Vordergrund gestanden haben, rückt nun der Mensch, der seiner Arbeit in einem mehr und mehr digitalen Umfeld nachgeht, in den Mittelpunkt der Diskussion. In Anlehnung an den Begriff *Industrie 4.0* hat sich der Begriff *Arbeit 4.0* [Bun17] etabliert. Durch die Verwandtschaft der Begrifflichkeiten wird betont, dass sich mit jeder der vier industriellen Revolutionen die Arbeitsbedingungen deutlich verändert haben [Alt+17]. Auch in den aktuellen Entwicklungen der Industrie kommt es vor allem zu einer Änderung der Produktionsarbeit. Zum einen steigt der Bedarf an immer individuelleren Produkten. Zum anderen werden zur Produktion genau dieser Produkte häufig menschliche Fähigkeiten benötigt, um ein zufriedenstellendes Resultat zu erhalten. Der Mensch wird jedoch schon bald nicht mehr in der Lage sein, sich sämtliche Variationen eines Produktes zu merken und reproduzieren zu können. Aus diesem Grund wird der Bedarf an effizienten und nutzerfreundlichen Assistenzsystemen weiter steigen. Um einen effizienten Ablauf der Produktion gewährleisten zu können, sind Assistenzsysteme nötig, die auf optimalen Prozessen basieren und somit die Beschäftigten bestmöglich unterstützen.

Zur Erstellung dieser Assistenzsysteme und der darin enthaltenen Montageprozesse wird häufig das in der Produkt- und Softwareentwicklung verbreitete Konzept der *nutzerorientierten Entwicklung* (engl. *User-driven Development*) angewendet. Dabei rücken die Personen in den Mittelpunkt des Entwicklungsprozesses, die schlussendlich das Endprodukt (in diesem Fall ein Montageassistenzsystem) nutzen. Sie werden bereits zu einem frühen Zeitpunkt in die Entwicklung einbezogen, um das Endprodukt bestmöglich zu ge-

stalten. Anhand ihren Einschätzungen und ihren Bewertungen wird das System verbessert, bis die nutzenden Personen hinreichend zufrieden sind.

Gerade für die Akzeptanz von Assistenzsystemen, die dauerhaft und permanent genutzt werden sollen, ist es besonders wichtig, dass die tatsächlich nutzenden Personen bei der Entwicklung des Systems mitwirken und ihre Erfahrungen und Wünsche einbringen können. Dabei wird in den meisten Fällen ein Prototyp des Systems verwendet und nach standardisierten Methoden bewertet.

Aus der menschlichen Verhaltens- und Interaktionsforschung ist bekannt, dass Menschen nonverbale Signale wie beispielsweise Mimik nutzen, um Informationen zu kommunizieren [WBJ67]. Dabei werden permanent und auch unterbewusst Signale gesendet. Diese Signale zeigt der Mensch sowohl bei der Interaktion mit anderen Menschen als auch bei der Interaktion mit künstlichen Systemen und Maschinen [Zen+09]. Menschen sind in der Lage, diese Signale eines Interaktionspartners zu lesen, zu interpretieren und auf sie zu reagieren. Auch Systeme sind bereits in der Lage menschliches Verhalten zu analysieren und in den Menschen *hineinzuschauen*.

In dieser Arbeit zeige ich, dass eine automatische Analyse von menschlichen Verhaltensweisen bei der Interaktion mit einem System als Evaluationskanal in einer nutzerorientierten Entwicklung verwendet werden kann. Diese neuartige Evaluation ist bei der Anwendung in prozessbasierten Systemen in der Lage durch eine konsequente automatisierte Analyse des Verhaltens der nutzenden Person verbesserungswürdige Prozessschritte aufzuzeigen.

Bislang wird in der nutzerorientierten Entwicklung auf eine subjektive Selbsteinschätzung der Versuchspersonen zurückgegriffen. Dabei erfolgt die Evaluation zumeist durch standardisierte Evaluationsmechanismen wie beispielsweise Fragebögen oder Interviews, die nach der Aufgabenerfüllung ausgefüllt und durchgeführt werden. Es kommt zu der Herausforderung, dass die befragte Person ausschließlich die Störungen benennen kann, die ihr noch bewusst sind bzw. die ihrer Meinung nach nennenswert sind. Gerade bei Prozessen, die lange dauern, wird eine Erinnerung an sämtliche Störungen schwierig und es steigt das Risiko, dass leichte Prozessbeeinträchtigungen nicht genannt werden. Außerdem schätzt die Person die Aufgabe individuell ein und kann ihre Anmerkungen bewusst beeinflussen, um beispielsweise der Norm zu genügen oder einen guten Eindruck zu machen. Mit der Verwendung der Verhaltensanalyse wird ein objektives Verfahren entwickelt, das eine Selbsteinschätzung zu weiten Teilen ersetzen kann.

Besonders in der Entwicklung von prozessbasierten Anwendungen wie der manuellen Montage bringt eine solche Evaluationsmöglichkeit einen großen Fortschritt. So werden gerade Prozesse, die der Einarbeitung und dem Anlernen von neuen Arbeitskräften dienen, von Experten erstellt [Agu04]. Diese Experten haben umfassende Erfahrung mit dem Prozess. Es besteht allerdings das Risiko, dass sie Annahmen treffen, die für erstmalige Benutzer so nicht getroffen werden können. Eine Analyse des Nutzungsverhaltens einer unerfahrenen Person bei der Prozessdurchführung deckt diese Schwierigkeiten auf und kann dem Prozessentwickler als Werkzeug zur optimalen Prozessentwicklung dienen.

In dieser Arbeit zeige ich anhand eines nutzerorientierten Entwicklungsprozesses am Beispiel eines Montageassistenzsystems, dass durch die Analyse menschlichen Verhaltens Verbesserungspotentiale im modellierten Prozess aufgedeckt werden, die Standardmechanismen der nutzerorientierten Entwicklung in dieser Art nicht erkennen. Dabei bietet die Einbettung der Verhaltensanalyse in der nutzerorientierten Systementwicklung folgende Vorteile:

- Optimierung des Assistenzsystems für eine zielgerichtete Hilfestellung der nutzenden Person
- Detaillierte Analyse des Prozesses und konkrete Aufdeckung problematischer Abschnitte
- Beschleunigung der Prototypenevaluationsphase durch Wegfall langwieriger Interviews und Fragebögen
- Beschleunigung der Prozessoptimierung durch konkrete Handlungsempfehlungen
- Objektive und kontinuierliche Analyse von Prozessen

1.1 Ziel dieser Arbeit

Ich verfolge in dieser Arbeit das Ziel, Verfahren zu entwickeln, um anhand menschlicher Verhaltensweisen während der Systembenutzung automatisiert prozessbasierte Systeme zu analysieren und so mögliche Schwachstellen und Potentiale in der Modellierung aufzudecken. Durch den hohen Bedarf an optimalen Assistenzsystemen und Montageprozessen in der industriellen Produktion werde ich die Verfahren durch eine System- und Prozessentwicklung in der manuellen industriellen Montage evaluieren und dadurch auch die Praxisnähe und die Nutzbarkeit zeigen.

In dieser Dissertation möchte ich zeigen, dass die menschliche Verhaltensanalyse für eine Prozessanalyse geeignet ist und die nutzerorientierte Entwicklung verbessern kann. Auf der Grundlage dieser Arbeit soll einem Prozessentwickler ein Werkzeug zur Verfügung gestellt werden können, das eine objektive Prozessanalyse anhand der direkten Beobachtung und Interpretation der prozessausführenden Person ermöglicht. Zusammenfassend werde ich in dieser Arbeit die folgenden Forschungsfragen betrachten:

Forschungsfrage 1 Wie kann in einem prozessbasierten Assistenzsystem eine multimodale menschliche Verhaltensanalyse von nutzenden Personen für eine Bewertung der einzelnen Prozessschritte verwendet werden?

Forschungsfrage 2 Können die entwickelten Methoden auf weitere Mensch-System Interaktionen übertragen werden?

1.2 Aufbau dieser Arbeit

Zu Beginn werde ich in Kapitel 2 (siehe Seite 5) die Grundlagen dieser Arbeit beschreiben. Dabei motiviere ich das gewählte Beispielszenario durch eine Charakterisierung der Rolle des Menschen in der industriellen Produktion. Darauf folgt eine Darstellung der nutzerorientierten Entwicklung, die ich im Verlauf dieser Arbeit anwende und deren Evaluationsphase durch die von mir entwickelten Verfahren erweitert wird. Das Kapitel endet mit einer Einführung der menschlichen Zustandserkennung anhand einer Gesichtsanalyse und der Vorstellung praxisnaher Anwendungen mit Bezug zu meiner Arbeit. In Kapitel 3 (siehe Seite 25) beschreibe ich den durchgeführten nutzerorientierten Entwicklungsprozess. Dieser dient als Grundlage und Basis der Datenerhebung für die entwickelten Verfahren zur Prozessschrittanalyse. Es folgt in Kapitel 4 (siehe Seite 39) eine Beschreibung des entwickelten Bildverarbeitungssystems zur Verhaltensanalyse und die Definition der im späteren Verlauf der Arbeit verwendeten Merkmalsgruppen. Darauf aufbauend werden in Kapitel 5 (siehe Seite 61) die vorgenommenen Anpassungen an dem entwickelten Assistenzsystem beschrieben, um eine standardisierte Evaluationsumgebung zu erhalten. Kapitel 6 (siehe Seite 77) und Kapitel 7 (siehe Seite 105) stellen die entwickelten maschinellen Lernansätze zur Prozessschrittbewertung dar. Zum Nachweis der Generalisierbarkeit werden diese Ansätze in Kapitel 8 (siehe Seite 121) auf ein Szenario der Bildungsforschung übertragen. Diese Arbeit schließt in Kapitel 9 (siehe Seite 133) mit einer Zusammenfassung, einem Fazit und einem Ausblick.

Hintergrund

” *One cannot not communicate.*

– Paul Watzlawick

Um die in dieser Arbeit entwickelten Ansätze einordnen zu können, erläutere ich in diesem Kapitel zunächst die Entwicklung der Digitalisierung und die sich ändernde Rolle des Menschen in der industriellen Produktion der Zukunft. Es folgt die Beschreibung der nutzerorientierten Entwicklung. Diese kann durch die von mir entwickelten Verfahren erweitert und komplementiert werden. Abschließend stelle ich die Gesichtsanalyse zur menschlichen Zustandserkennung vor und diskutiere bereits vorhandene praxisnahe Systeme der automatischen Interpretation menschlichen Verhaltens auf Basis einer Analyse des Gesichtes.

2.1 Der Mensch in der industriellen Produktion der Zukunft

Zurzeit befindet sich die Industrie und dadurch auch die Arbeit in der Industrie in einer Phase des Wandels. Hirsch-Kreinsen [Hir15a] spricht von einem „Megatrend der zukünftigen gesellschaftlichen Entwicklung“. Dabei ist die Einteilung in einzelne *Revolutionen* besonders in politischen Diskussionen weit verbreitet [Alt+17; BH15; Pro13]. Es ist jedoch ungewöhnlich, dass eine Revolution bereits im Vorfeld angekündigt und nicht wie üblich im Nachhinein durch historische Betrachtungen als solche identifiziert wird. Nur eine solche rückblickende Betrachtung kann eine *nachhaltige* und *abrupte* Umwälzung gesellschaftlicher Strukturen nachweisen. Diese gesellschaftlichen Änderungen sind bei den vergangenen Umwälzungen besonders im Hinblick auf die Arbeitsbedingungen zu sehen [Röb17]. Ob sich die Gesellschaft durch die aktuellen Entwicklungen nach der *Mechanisierung* (erste industrielle Revolution), *Elektrifizierung* (zweite industrielle Revolution) und

Automatisierung (dritte industrielle Revolution) erneut umwälzt oder ob wir uns noch immer in der dritten Revolution befinden [Pro13] kann nur die Zukunft zeigen. Röben [Röb17, S. 264] kann beispielsweise in den aktuellen Entwicklungen keine „wirkliche Revolution“ feststellen. Es sei wenig von einem „Hinwegfegen und Vernichten“ ganzer Teile der Industrie zu spüren. Als Auslöser der vierten industriellen Revolution wird dabei vor allem die *Vernetzung* von digitalen Systemen durch das *Internet der Dinge* [FM05; UHM11] und der *Cyber-physischen Systeme* [BG11] gesehen.

Egal, ob sich die Industrie gerade in einer Revolution befindet oder nicht, ist zu erkennen, dass sich industrielle Produktionsstätten deutlich verändern. Die physische Welt verschmilzt mit der virtuellen Welt und durch minimalisierte Computer können immer mehr physische Geräte untereinander kommunizieren. Durch diese umfassende Vernetzung wird die Umgebung des Menschen intelligenter und digitaler. Immer mehr Daten können gesammelt und ausgewertet werden [LRU14]. In der Industrie wurde dabei zunächst der Schwerpunkt auf Analysen der Produktionsanlagen und somit auf eine Weiterentwicklung der Automatisierung gelegt.

Diese zunehmende Digitalisierung bietet die Chance auf ein sich veränderndes Kundenverhalten zu reagieren. Der Bedarf an individuelle und immer speziellere Produkte wächst stetig. Exemplarisch sei hier die Konfigurationsvielfalt von Automodellen genannt. Um diese neuen Anforderungen in der Produktion bewältigen zu können, wurden ganzheitliche und digitale Konzepte für den gesamten Produktionsprozess entwickelt und bereits in vielen Unternehmen implementiert [Dum+15]. Dieser Prozess beginnt bereits bei der Bestellung und führt über die Produktion bis zur Montage und Auslieferung.

Dabei stand bislang häufig der Versuch der vollständigen Automatisierung im Vordergrund. Es wurde jedoch deutlich, dass eine Automatisierung in vielen Bereichen entweder aus wirtschaftlichen oder aus technischen Gründen nicht sinnvoll ist. Darüber hinaus ist in den letzten Jahren sogar die Entwicklung von der Automatisierung zurück zur menschlichen Arbeitskraft zu erkennen [Lux16; Ryz18].

Um dennoch die geänderten Anforderungen bewältigen zu können, rückt der Mensch, der in diesem digitalen Umfeld arbeitet, wieder in den Fokus der Diskussion. Die steigende Komplexität und Variantenvielfalt eines Produktes stellen aber auch Beschäftigte vor große Herausforderungen. Jede Person in der Produktion muss mit ihren individuellen Voraussetzungen (kognitive Fähigkeiten, körperliche Einschränkungen usw.) die immer komplexeren und individuelleren Produktionsprozesse bewältigen. Eine Effizienz wie bei der klassischen Fließbandarbeit kann dabei nicht erhalten werden.

Es stoßen also sowohl die Automatisierung als auch die manuelle Montage im herkömmlichen Sinne an ihre Grenzen und alternative Wege müssen gefunden werden. Der Bedarf wurde bereits erkannt und wird nicht zuletzt durch Einführung neuer Begrifflichkeiten wie *Human-in-the-Loop Cyber-Physical-Systems* [NBS17], *Cyber-Physical-Human Systems* [Sow+16], *Human Cyber-Physical Systems* [Rom+16] oder *sozio-technische Systeme* [Hir14] deutlich. Auch der Wissenschaftliche Beirat der Plattform Industrie 4.0 [Wis14] thematisiert explizit den Menschen in einer intelligenten Produktion und betont vor allem die „vielfältigen Möglichkeiten für eine humanorientierte Gestaltung der Arbeitsorganisation“.

Grundsätzlich werden in der aktuellen Forschung zwei gegensätzliche Perspektiven der Arbeit in Produktionsumgebungen der Zukunft aufgezeigt. Einerseits werden die positiven, andererseits die negativen Entwicklungsmöglichkeiten betont. Hierbei ist die Beschreibung des Positiven durch den Begriff *Werkzeugszenario* und des Negativen durch den Begriff *Automatisierungsszenario* weit verbreitet [Win14]. Bei diesen Begrifflichkeiten wird der Fokus speziell auf die Entscheidungsfreiheit der Mitarbeitenden gelegt. Auch ich werde diese Begrifflichkeiten verwenden und im Folgenden weiter erläutern.

2.1.1 Automatisierungsszenario

Das Automatisierungsszenario beschreibt eine Produktion der Zukunft, in welcher die Beschäftigten die Kontrolle über die Abläufe vollständig abgeben. Dabei würde es zu „enorm[n] Substitutionseffekte[n]“ und „verschärfte[n] Kontroll- und Anweisungsstrukturen“ kommen [IN15]. Die Arbeitskräfte wären rein ausführende Organe und hätten die Entscheidungsfindung den automatisierten Systemen übergeben. Es würden nur die Aufgaben, die noch nicht automatisiert werden konnten, beim Menschen verbleiben [Deu+15].

Dieses Szenario birgt große Gefahren, die als *Ironien der Automatisierung* bereits in den 1980er Jahren beschrieben wurden [Bai83; Lüd15]:

- Ironie 1** Menschen werden als fehleranfällig angesehen und werden aus den Prozessen durch Entwickler entfernt. Der Entwickler selbst ist aber auch ein Mensch und genauso anfällig für Fehler.
- Ironie 2** Nicht automatisierbare und somit komplexe Aufgaben werden auf den Menschen übertragen, wobei dieser doch eigentlich als zu substituierendes schwächstes Glied im Prozess angesehen wird.

- Ironie 3** Nach der Substitution des Menschen durch die Maschine soll dieser aber dennoch die Maschinen überwachen und bei Fehlern eingreifen.
- Ironie 4** Robuste automatisierte Systeme erfordern viel Training der Mitarbeitenden, da diese sich nicht permanent mit dem System auseinandersetzen müssen und somit die Abläufe nicht durch alltägliche Arbeit verinnerlicht haben. Instabile automatisierte Systeme erfordern hingegen durch ständiges Eingreifen des Menschen und damit verbundenem besseren Verständnisses des Systems weniger Training.

Die Ironien wurden bereits in der Vergangenheit bei der Einführung autonomer Systeme diskutiert, um mögliche Gefahren aufzudecken. Gefahrensituationen sind unter anderem bei der Automatisierung der Steuerung von Flugzeugen aufgetreten [Lüd15], werden zurzeit bei der Entwicklung autonomer Autos und LKW diskutiert [MJ17; CHN16] und können auf industrielle Produktionen übertragen werden. Auch dort wurde bislang angestrebt, den Menschen als schwächstes Glied im Prozess durch technische Systeme zu ersetzen. Dennoch soll er, wenn die Automation an ihre Grenzen stößt, die aktuelle Umgebungssituation schnell und vor allem richtig einschätzen und adäquat reagieren, um die Produktion aufrechtzuerhalten. Der Mensch soll in der Lage sein, richtig und schnell zu reagieren, wenn vollständig autonome Maschinen nicht mehr funktionieren oder ausfallen. Durch den hohen Automatisierungsgrad verliert der Mensch jedoch das Verständnis für den Prozess und kann bei Unstimmigkeiten nicht mehr angemessen handeln. Somit kann er in Problemfällen auch nicht mehr reparierend eingreifen.

2.1.2 Werkzeugszenario

Es könnte durch die aktuellen Entwicklungen aber auch zu neuen „Beschäftigungsperspektiven“ und „Freiheiten im Arbeitsprozess“ kommen [IN15]. Diese werden als Werkzeugszenario beschrieben. Automatisierte Systeme werden dabei als Werkzeuge betrachtet, die den Menschen unterstützen. Durch transparente und datengetriebene Produktionssysteme könnten die Menschen optimal und gewinnbringend unterstützt werden. Romero u. a. [Rom+16] beschreiben das Werkzeugszenario als *Operator 4.0* bzw. *Work aided by machines*. Lüdtker [Lüd15], Christoffersen und Woods [CW02] und Klein u. a. [Kle+04] nennen die

enge Zusammenarbeit zwischen Mensch und Maschine *Mensch-Maschine Teams*. Die Systeme würden Auswertungen und Analysen liefern, um den Menschen bei Entscheidungen zu unterstützen. Die finale Entscheidung träfe aber der Mensch und kein System. Nach Kagermann [Kag17] sei das Werkzeugszenario eine *soziale Chance*. Die Produktion würde dem Takt des Menschen folgen. Mitarbeiter würden nicht als reine Maschinenbediener eingesetzt, sondern in der Rolle des *Erfahrungsträgers*, *Entscheiders* und *Koordinators* rücken [Kag17]. Darüber hinaus stellen Töniges u. a. [Tön+16] fest, dass eine Schlüsselrolle für gute Arbeitsbedingungen die richtige Balance zwischen Unterstützung und Autonomie des Menschen sei und dass dem Menschen die Kontrolle über seine Arbeit nicht genommen werden dürfe.

2.1.3 Diskussion

Welche der beiden Entwicklungsrichtungen in Zukunft überwiegen wird, kann nicht vorhergesagt werden. Es wird von *Möglichkeiten* [IN15], *Gestaltungsalternativen* [Hir15a], oder *Entwicklungsperspektiven* [Hir15b] gesprochen. Den Entscheidungsträgern in den einzelnen Unternehmen muss jedoch klar sein, dass die Entscheidung über die Zukunft der Arbeit in der Produktion bewusst getroffen werden muss. Es sind durchaus Mischformen beider beschriebener Szenarien möglich, die jeweilige Umsetzung muss aber bewusst eingeleitet und vorangetrieben werden.

Die von mir in dieser Arbeit vorgestellten Verfahren zur Verbesserung von Systemen und Prozessen sind klar im Bereich des Werkzeugszenarios einzuordnen. Durch die Einführung eines technischen Systems wird der System- und Prozessentwickler unterstützt und auf mögliche Verbesserungspotentiale hingewiesen. Eine autonome Anpassung durch das System erfolgt nicht. Durch die konsequente Umsetzung als Werkzeug kann eine technische Umgebung geschaffen werden, die die Entscheidungsgewalt über Prozessanpassungen beim Menschen belässt. Dem Menschen werden objektive Informationen über den Prozess zur Verfügung gestellt. Er kann diese Information nutzen, um Prozesse besser bewerten und eine fundiertere Entscheidung über Änderungen treffen zu können.

Ich bin der Meinung, dass ein flächendeckendes Umdenken und die konsequente Implementierung von intelligenten technischen Systemen als Werkzeuge zur Unterstützung des Menschen zu einer erneuten Umwälzung der Industrie führen können. Dieses kann sich als die wahre vierte industrielle Revolution herausstellen. Durch eine Unterstützung des Menschen durch technische Systeme wird dieser in der Lage sein, die beschriebene

steigende Komplexität der Produktionsprozesse zu bewältigen. Somit kann eine effiziente Produktion auch in Zukunft gewährleistet werden.

2.2 Nutzerorientierte Entwicklung

In der nutzerorientierten Entwicklung steht die letztendlich nutzende Person eines Produktes, eines Systems oder auch eines Prozesses im Mittelpunkt. Diese Art der Entwicklung ist auch unter den Begriffen *User-driven Development*, *User-centered Design*, *Human-centered Design* oder *Nutzerorientierte Gestaltung* bekannt. Norman und Draper [ND86] haben diese Entwicklung erstmals beschrieben. Mit der Zeit wurde sie auf unterschiedlichste Szenarien angepasst [Nor13; AMP04]. Nutzerorientierte Systementwicklung führt in vielen Fällen zu einer Verbesserung der Nützlichkeit und Benutzbarkeit von Produkten [Vre+02]. Die Wichtigkeit der Zentrierung auf die nutzenden Personen wird nicht zuletzt durch die diesbezüglich definierte ISO Norm 9241–210 [Int10] zum nutzerorientierten Design von interaktiven Systemen deutlich. Seine Ursprünge hat die nutzerorientierte Entwicklung in der Produktentwicklung. Erst später wurden die Konzepte auch auf die Softwareentwicklung ausgeweitet. Die nutzenden Personen werden in sämtliche Schritte des Entwicklungsprozesses einbezogen, um ein maßgeschneidertes Endresultat zu garantieren.

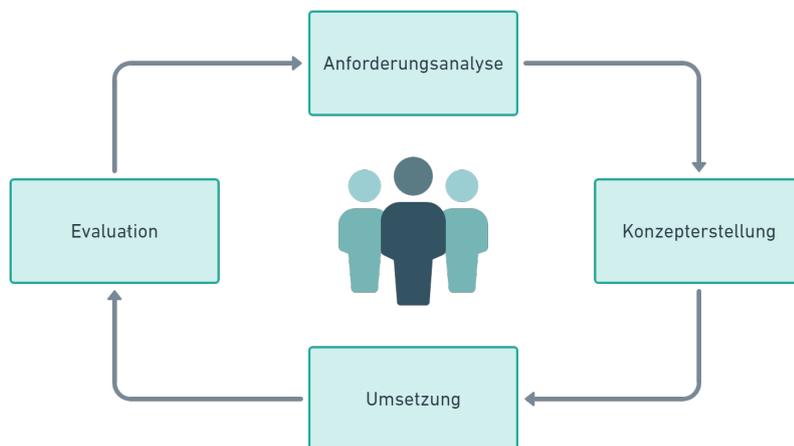


Abbildung 2.1.: Nutzerorientierter Entwicklungsprozess:

Es wird die Anforderungsanalyse, die Konzepterstellung, die Umsetzung und die Evaluation solange iteriert, bis die Anforderungen der nutzenden Personen hinreichend erfüllt sind [ND86].

In Abbildung 2.1 ist das Prinzip der nutzerorientierten Entwicklung zu sehen. Zunächst werden in der Anforderungsanalyse die Bedürfnisse der nutzenden Personen strukturiert aufgenommen und analysiert. Dieses geschieht beispielsweise durch standardisierte Interviews oder Fragebögen. Ergebnisse dieser Phase sind zum Beispiel *User Stories* [Coh04] oder *Use Cases* [ACB02]. In der darauffolgenden Konzepterstellung werden erste Prototypen auf Basis der vorangegangenen Analyse erstellt. Auch diese Prototypen werden gemeinsam mit den nutzenden Personen erstellt und durch diese evaluiert. In der dritten Phase folgt die konkrete technische Umsetzung, in der das Produkt mithilfe der zuvor gewonnenen Erkenntnisse implementiert wird. Es folgt die Evaluation des umgesetzten Systems durch die nutzenden Personen. Abschließend lässt sich der Kreis schließen und eine neue Iteration beginnen, indem die Ergebnisse der Evaluation und die definierten Anforderungen in einer Analysephase miteinander in Verbindung gebracht werden und auf dieser Basis der Kreislauf erneut durchlaufen wird. Dieses erfolgt solange, bis die Optimierung abgeschlossen und ein für die nutzenden Personen optimales Produkt entstanden ist.

Aus einer erfolgreichen nutzerorientierten Entwicklung folgt zumeist ein gutes *Nutzungserlebnis* des Produktes (engl. *User Experience*). Dieser Begriff wurde auch durch Norman u. a. [NMH95] erstmalig genannt und geprägt. Dabei wird vor allem die Wichtigkeit der Emotionalität und affektiver Faktoren der nutzenden Personen bei der Benutzung von Systemen herausgestellt. Sie müssten sich bei der Systembenutzung „wohl fühlen“. Außerdem ist das Nutzungserlebnis eine ganzheitliche Beschreibung, die sowohl die Emotionen [HT06], die psychologischen und physiologischen Reaktionen, als auch Erwartungen, Vorkenntnisse und allgemeine Eigenschaften des Nutzers einbezieht.

In der Beschreibung der nutzerorientierten Entwicklung und des optimalen Nutzungserlebnisses werden Evaluationsmöglichkeiten genannt, die entwickelte Prototypen bewertbar machen. Im folgenden Abschnitt werde ich die wichtigsten und für diese Arbeit relevanten Evaluationsmöglichkeiten vorstellen, um die entwickelten Ansätze einordnen zu können.

2.2.1 Evaluationsmöglichkeiten

Selbstbeurteilung

Durch standardisierte und validierte Umfragen werden subjektive Empfindungen nach der Bearbeitung einer Aufgabe mit dem System quantifiziert und vergleichbar gemacht. Zur direkten Erhebung des Nutzungserlebnisses wird beispielsweise der *User Experience Questionnaire (UEQ)* angewendet [LHS08]. Darüber hinaus werden oftmals Fragebögen aus der kognitiven Belastungsforschung wie beispielsweise der *NASA-Task Load Index (NASA-TLX)* [HS88] oder die *Subjective Workload Assessment Technique (SWAT)* [RN88] verwendet. Weitere Informationen zur Ableitung und Erhebung menschlicher Belastung fasst Cain [Cai07] zusammen. Eine allgemeingültige und einheitliche Definition von kognitiver Belastung kann jedoch nicht gefunden werden. Dieses mag vor allem daran liegen, dass die zugrundeliegenden psychologischen Prozesse noch nicht vollständig verstanden sind [Cai07]. Es wird jedoch angenommen, dass durch eine Steigerung der kognitiven Belastung die Leistung abnimmt, die Reaktionszeit zunimmt und auch die Fehlerrate steigt [Cai07]. Bei einer Übertragung auf manuelle Montagetätigkeiten führt somit eine hohe kognitive Belastung zu einer schlechteren Qualität des Produktes. Die kognitive Belastung sollte also auch in der Arbeitsgestaltung berücksichtigt werden, um eine hohe Qualität zu gewährleisten. Schon Casali und Wierwille [CW84] haben Belastung als ein Konstrukt beschrieben, das nicht direkt beobachtet werden könne, sondern von Verhaltensweisen bzw. psychologischen und physiologischen Mustern abgeleitet werden müsse. Joiko u. a. [JSW02] beschreiben darüber hinaus die Wichtigkeit einer psychischen Belastungsermittlung in Arbeitsprozessen. In dieser Arbeit wird mit dem *Raw Task Load Index (RTLX)* [Har06] eine adaptierte Version des NASA-TLX verwendet. Dabei wird entgegen dem ursprünglichen Fragebogen auf eine Gewichtung der einzelnen Dimensionen verzichtet. Der RTLX liefert vergleichbare Ergebnisse und ist deutlich einfacher zu erheben [Har06].

Annett [Ann02] stellt heraus, dass eine Selbsteinschätzung der Belastung rein subjektiv und eine Generalisierbarkeit über unterschiedliche Personen schwer möglich sei. Jede Person hätte eine individuelle Einschätzung ihrer Belastung und könne diese nicht absolut einordnen. Auch Cain [Cai07] beschreibt, dass oftmals die eigene Leistung falsch eingeschätzt würde und somit die eigene Beurteilung nicht auf objektiven Gegebenheiten beruhe. Diese könne nur einen groben Anhaltspunkt liefern [MH88]. Außerdem muss der Nutzer die zu bewertende Aufgabe rückblickend einschätzen. Je nach Dauer der Aufgabe kann die Erinnerung bereits verblasen oder es können auch Anmerkungen vergessen

bzw. durch stärkere Eindrücke verdrängt worden sein. Kontinuierliche und objektive Bewertungen sind mithilfe von Selbstbeurteilungsmethoden ohne Beeinträchtigung der eigentlichen Aufgabe nicht möglich.

Systembeurteilung

Bei der Systembeurteilung (engl. *Usability Tests*) wird im Gegensatz zur Selbstbeurteilung das System und dessen Funktionalität bewertet. Beispielsweise wird evaluiert, wie gut oder wie schnell eine Aufgabe mit dem System gelöst werden kann. Dabei nimmt die Emotionalität der nutzenden Person eine untergeordnete Rolle ein. Bei der Systembeurteilung werden auch zumeist standardisierte Fragebögen verwendet. Brooke u. a. [Bro+96] haben einen weit verbreiteten Fragebogen entwickelt, mit dessen Hilfe der *System Usability Score (SUS)* berechnet werden kann. Es werden 10 Fragen mit jeweils einer Wertungsskala von 1 (stark ablehnend) bis 5 (stark zustimmend) erhoben. Daraus wird der SUS (0–100) berechnet mit dem auf die Benotung (A–F) der Systemnutzbarkeit geschlossen werden kann. Auch in dieser Arbeit erhebe ich zur Evaluierung des entwickelten Montageassistenzsystems den SUS. Eine ähnliche Kennzahl ist die *Usability Metric for User Experience (UMUE)* [Fin10]. Der zugrundeliegende Fragebogen besteht aus vier Fragen mit einer jeweiligen Skala von 1 (stark ablehnend) bis 7 (stark zustimmend). Der *Net Promoter Score (NPS)* [Rei03] reduziert den Fragenkatalog weiter und stellt nur eine einzelne Frage („Würden Sie das Produkt einem Freund oder Kollegen empfehlen?“). Es ist zu beachten, dass auch Systembeurteilungen reine subjektive Einschätzungen der einzelnen Probanden sind.

Aufgabendauer

Besonders in der Prozessmodellierung und -optimierung von manuellen Montageprozessen werden Konzepte zur Abschätzung der zu erwartenden Dauer eines einzelnen Montageschrittes verwendet. Dabei wird auch die Gestaltung des konkreten Arbeitsplatzes einbezogen und optimiert. Korn [Kor14, Kapitel 2.6.1] fasst diese Modellierung zusammen. Hierbei ist besonders die *Arbeitsablauf-Zeitanalyse (AAZ)* (engl. *Methods-Time Measurement (MTM)*) ein weit verbreitetes Konzept, um die Güte von Prozessen anhand einer zeitlichen Analyse aufzuzeigen. Menschliche Handlungs- und Bewegungsabläufe werden mit empirisch erhobenen Zeiten quantifizierbar gemacht. Diese Richtzeiten definieren, wie lange einzelnen Abläufe im Schnitt dauern. Mit jeder Aktion ist eine Dauer assoziiert, die in *Time Measurement Units (TMUs)* angegeben wird.

Die AAZ ist standardisiert und wird weltweit eingesetzt. In Deutschland wird sie vom *Verband für Arbeitsgestaltung, Betriebsorganisation und Unternehmensentwicklung (REFA)* stetig weiterentwickelt. Prozesse und Arbeitsplätze sind dabei optimal, wenn die zu lösenden Aufgaben eine möglichst geringe Dauer zur Durchführung benötigen.

Diese Modellierung ist eine direkte Abschätzung der Güte eines Prozesses. Der Prozess wird direkt analysiert und die einzelne Person muss nicht befragt werden. Zur Modellierung und Entwicklung eines Arbeitsplatzes hat eine Optimierung im Hinblick auf schnelle Arbeitsabläufe durchaus eine Berechtigung und wird in vielen Produktionen eingesetzt. Logische Fehler (beispielsweise in der Prozessschrittbeschreibung) können durch diese Methode aber nicht aufgedeckt werden.

Die Dauer einzelner Prozessschritte kann auch direkt verwendet werden, um den jeweiligen Schritt empirisch im Vergleich zu anderen Prozessschritten zu bewerten. Falls in den Schritten ähnliche Aufgaben erfüllt werden, kann die zeitliche Analyse aufdecken, ob Probleme in einzelnen Aufgabenschritten entstanden sind.

Systembedienverhaltenanalyse

Auch das Bedienverhalten des Systems durch die nutzenden Personen kann Rückschlüsse auf die Güte des Systems und des zugrundeliegenden Prozesses liefern. Diese Analyse ist wie eine zeitliche Analyse direkt ableitbar. Wird beispielsweise in einem Prozessschritt häufig die Hilfe aufgerufen oder wird ein Bild durch viele Personen vergrößert, kann dieses ein Zeichen für einen verbesserungswürdigen Prozessschritt sein. Auch das in dieser Arbeit entwickelte System berücksichtigt eine Analyse des Bedienverhaltens zur Bewertung der Prozessschritte.

Augenanalyse

Ein weiteres direktes Evaluationsverfahren basiert auf einer Analyse der Augen der nutzenden Personen. Dabei können beispielsweise Blicke analysiert werden, um Aussagen über das System abzuleiten [GK13].

Darüber hinaus kann die Pupillengröße als Merkmal verwendet werden (*Pupillometrie*). Durch eine Änderung der Pupillengröße (unter gleichbleibenden Umgebungsbedingungen)

kann eine Änderung der kognitiven Belastung erkannt werden. Detaillierte Zusammenfassungen der Forschung in diesem Bereich wurden beispielsweise durch Szulewski u. a. [SKH17] oder Laeng u. a. [LSG12] erstellt.

Die Augenanalyse ist ein direkter Kanal, der innere Zustandsänderungen des Menschen ableitbar macht und Informationen über den aktuellen Zustand der Versuchsperson liefern kann. Dabei müssen aber besonders für eine Anwendung der Pupillometrie die Umgebungsbedingungen (u. a. Beleuchtung) konstant gehalten werden. Ein Einsatz in unkontrollierbaren Produktionsumgebungen ist mit den aktuellen Methoden nicht zu realisieren und wird somit auch in dieser Arbeit nicht verwendet. Eine Analyse der Blickrichtung wird jedoch auch von mir benutzt, um weitere Informationen über das Nutzungsverhalten extrahieren zu können.

Vitalparameteranalyse

Um Vitalparameter (beispielsweise Herzfrequenz, Herzratenvariabilität, Blutdruck oder Atemfrequenz) ableitbar zu machen, kommt in den meisten Fällen körpernahe Sensorik zum Einsatz. Dabei werden Sensoren direkt auf der Haut angebracht und ermöglichen eine direkte Analyse. Vitalparameter lassen des Weiteren Rückschlüsse auf den mentalen Zustand der Nutzenden zu.

Herzratenvariabilität kann beispielsweise mit kognitiver Belastung in Verbindung gebracht werden. Dieses wurde bereits bei alltäglichen Büroaufgaben [Cin+13], beim Vergleich verschiedener Automatisierungsgrade [Du+15] oder bei kognitiven Herausforderungen [Hör+16] gezeigt.

Du u. a. [Du+15] stellen fest, dass auf der einen Seite die subjektive Wahrnehmung von Belastung durch die Automation abnimmt (Erhebung durch NASA-TLX Fragebogen), die körperlichen Merkmale auf der anderen Seite aber für eine Zunahme der Belastung sprechen. Dieses unterstützt die bereits genannte Kritik an subjektiver Selbstbeurteilung und die Wichtigkeit weiterer Merkmalsanalysen wird deutlich.

Auch Hirnaktivität kann mit kognitiver Belastung, Müdigkeit, Erschöpfung und Aufmerksamkeit in Verbindung gebracht werden. Dieses wurde bereits bei der Analyse von Piloten [Bor+14], Autofahrern [Bor+14], in Flugsimulationsumgebungen [Dur+14] und in Montageassistenzsystemen [Mij+16] gezeigt.

Die beschriebenen Ergebnisse der Analyse vitaler Parameter des Menschen durch körpernahe Sensorik sind vielversprechend und es wurden auch bereits Anwendungen dieser Analyse in industriellen Montageumgebungen vorgestellt [Mij+16]. Dennoch verzichte ich auf die Berücksichtigung dieser Kanäle, da eine Verwendung nur durch eine komplexe Einrichtung und Sensorik direkt am Körper der Versuchspersonen möglich ist. In dieser Arbeit konzentriere ich mich auf die Verwendung eines Kamerasystems (siehe Abschnitt 5.1.1, Seite 62) und darauf basierenden Verfahren ohne langwierige Einrichtung. Dadurch kann ich eine möglichst hohe Praxistauglichkeit erzielen.

2.2.2 Diskussion

Nutzerorientierte Entwicklung ist in der Produkt-, der Software- und auch in der Assistenzsystementwicklung ein verbreitetes Konzept, das häufig zur Anwendung kommt. Ich habe besonders die vorhandenen Evaluationsmöglichkeiten dargestellt, die Verwendungsmöglichkeiten im Rahmen dieser Arbeit beschrieben und die Methoden kritisch hinterfragt. Es fällt auf, dass weit verbreitete Methoden auf subjektiven Einschätzungen und nicht auf einer direkten Analyse von Informationen beruhen. Im Falle von direkten und kontinuierlichen Methoden kann es zu der Verwendung von spezieller Sensorik kommen, die der Nutzer am Körper tragen muss. Solche Sensorik wird in den von mir vorgestellten und entwickelten Verfahren nicht benötigt, da eine handelsübliche Kamera zur Analyse des Kopfes ausreicht. Alle weiteren von mir verwendeten Merkmale können aus der Aufzeichnung der direkten Systemeingaben der nutzenden Person extrahiert werden.

Auch das Konzept der nutzerorientierten Entwicklung wird kritisiert. So formuliert Norman [Nor05] selbst in seinem Artikel, dass durch die Fokussierung auf die nutzende Person keine neuen und innovativen Produkte entstehen würden, da nicht die Zukunft eines Produktes, sondern die aktuellen Bedürfnisse der Nutzenden betrachtet würden. Dieses trifft auf das in dieser Arbeit betrachtete Szenario nicht zu, da kein neues Produkt entwickelt werden muss, sondern die Funktionsweisen des zu entwickelnden Assistenzsystems bereits bekannt sind. Außerdem stellt Norman [Nor05] heraus, dass Meinungen einzelner Personen oftmals so verschieden seien, dass sie nicht zusammengefasst werden können. Durch die in dieser Arbeit vorgestellten Verfahren wird direkt das menschliche Verhalten analysiert. Natürlich können sich auch Menschen in denselben Situationen (wie beispielsweise einem Prozessschritt) unterschiedlich verhalten. Die beschriebenen Verfahren sollten in der Lage sein mit diesen Situationen umzugehen und diese hinreichend zu

generalisieren. Die Gefahr ist dennoch gegeben, dass sich menschliche Verhaltensweisen bei der Durchführung von Aufgaben nicht generalisieren lassen und dadurch kein gutes Ergebnis in einer automatischen Analyse generiert werden kann. Aus diesem Grund wird auch die Generalisierbarkeit der Ansätze und der Methodentransfer auf andere Bereiche in dieser Arbeit betrachtet und hinterfragt.

Die von mir im Laufe dieser Arbeit vorgestellten Verfahren können als neue Evaluationsmethode der nutzerorientierten Entwicklung verwendet werden. Die Verwendung der Gesichtsanalyse als Basis dieser Verfahren bietet die Möglichkeit, eine objektive und verhaltensbasierte Analyse der nutzenden Person zur Bewertung von Systemen durchzuführen. Dabei können die Vorteile beispielsweise der Vitalparameteranalyse (Objektivität, Kontinuität) genutzt und die Nachteile der körpernahen Sensorik durch die Verwendung einer handelsüblichen Kamera vermieden werden.

2.3 Gesichtsanalyse zur menschlichen Zustandserkennung

Ich habe bereits im vorherigen Abschnitt 2.2 beschrieben, dass ein gutes Nutzungserlebnis eines Systems auf Emotionalität und somit auf den inneren Zustand der nutzenden Person beruht [HT06]. Die Empfindungen während der Interaktion sind meist nur schwer verbalisierbar, spielen aber eine wichtige Rolle in der Bewertung des Systems. Eine direkte und automatische Bewertung des Nutzungserlebnisses anhand dieser inneren menschlichen Zustände wäre also wünschenswert. In diesem Kapitel zeige ich, dass gerade diese Emotionalität und inneren Zustände anhand menschlicher Verhaltensweisen und besonders anhand der Mimik geschätzt werden können. Damit motiviere ich die Verwendung der Gesichtsanalyse als Kernmethodik der von mir entwickelten Verfahren.

Das menschliche Gesicht ist ein wichtiger, wenn nicht sogar der wichtigste Kanal der *nonverbalen Kommunikation* [WBJ67]. Es ist allgemein bekannt, dass der Mensch neben der verbalen Kommunikation in sozialen Interaktionen weitere interpretierbare Signale aussendet. Neben der Mimik gehören hierzu beispielsweise auch Gestik oder Körperhaltung. Nonverbale Kommunikation ist oftmals nicht durch den Menschen beeinflussbar (beispielsweise Erröten). Bestimmte Signale werden darüber hinaus unbewusst gesendet und der jeweiligen Person erst bewusst, wenn man sie darauf hinweist. Die Mimik kann der

Mensch zwar bewusst beeinflussen, unbewusste Änderungen können aber dennoch auch im Gesicht auftreten. Ein Beispiel für eine bewusste Beeinflussung ist das *soziale Lächeln*. Hierbei sind bereits Säuglinge in der Lage Mimik einzusetzen, um auf ihre Umgebung zu reagieren und diese zu beeinflussen [MF07]. Ein Beispiel für eine unbewusstes Signal ist die *Mikromimik*. Diese Gesichtsausdrücke haben eine sehr kurze Dauer und können kaum unterdrückt werden [HI66].

Fridlund [Fri91] stellte die These auf, dass der soziale Kontext einen höheren Einfluss auf die Mimik eines Menschen hätte als der innere Zustand. Diese These wurde durch Hess u. a. [HBK95] und Jakobs u. a. [JMF99] jedoch entkräftet. Es lässt sich aber dennoch festhalten, dass die Mimik durch zwei Kanäle beeinflusst wird. Auf der einen Seite wird diese durch den gegenwärtigen inneren Zustand, auf der anderen Seite durch das soziale Umfeld und den sozialen Kontext bedingt.

Welcher Kanal genau, welche Änderungen im Umfeld oder welcher innere Zustand eine Anpassung der Mimik hervorgerufen hat, ist jedoch in dem hier betrachteten Szenario nicht wichtig. Für eine Bewertung des Systems ist vielmehr essenziell, dass überhaupt eine Änderung bei der Interaktion mit einem System aufgetreten ist. Welchen Ursprung diese Änderung hat ist zweitrangig. Diesem Standpunkt folgend stellen auch Vinciarelli u. a. [VPB09] und Hudlicka [Hud03] heraus, dass eine Erkennung und Interpretation *sozialer Signale* die Mensch-Maschine Interaktion verbessern kann. Auch Lang u. a. [Lan+12] stellen die Wichtigkeit einer Analyse von *Facial Communicative Signals* in der Mensch-Roboter Interaktion dar und fokussieren sich in ihrer Arbeit auch auf die Analyse des Gesichtes. Verschiedene Verfahren zur Analyse dieser Signale werden neben Vinciarelli u. a. [VPB09] auch durch Tao und Tan [TT05] und Picard u. a. [PVH01] beschrieben und zusammengefasst. Martinez u. a. [Mar+17] sowie Cohn und La Torre [CL14] konzentrieren sich in ihren Übersichtsartikeln vor allem auf unterschiedliche Arten der Gesichtsanalyse.

2.3.1 Formale Beschreibung der Mimik

Mimik wird durch die Aktivität einer Vielzahl von Muskeln im Gesicht erzeugt. Dabei sind die Bewegungen im Gesicht und somit die Gesichtsausdrücke eines Menschen endlich und für jeden Menschen gleich. Diese endlichen Gesichtsausdrücke lassen sich systematisch beschreiben. Dabei ist das *Facial Action Coding System (FACS)* von Friesen und Ekman [FE78] als Beschreibungssprache weit verbreitet. Jede mögliche Bewegung, die durch Muskelkontraktion im Gesicht ausgelöst wird, lässt sich durch eine *Action Unit*

(AU) beschreiben. Es sind 44 unterschiedliche AUs definiert, die jeweils eine unabhängige Bewegung im Gesicht referenzieren. In Tabelle 4.1 (siehe Seite 47) sind die AUs beschrieben, die in dieser Arbeit verwendet werden. AUs treten zumeist in Kombination miteinander auf und können verschiedene Ausprägungsintensitäten haben (A (gerade erkennbar) – E (physiologisches Höchstmaß)). Auch Kopfbewegungen können durch das FACS abgebildet werden.

Ekman u. a. [EFA80] beschreiben durch die Kombination unterschiedlicher AUs universelle Basisemotionen (Wut, Freude, Trauer, Ekel, Verachtung, Angst, Überraschung). Beispielsweise sei gleichzeitiges Auftreten von AU 6 und AU 12 ein Zeichen für Freude oder gleichzeitiges Auftreten von AU 1, AU 2, AU 5 und AU 26 ein Zeichen für Überraschung. Somit ist die Verbindung der reinen Beschreibung des Gesichtsausdrucks mit dem inneren Zustand des Menschen und seinen Emotionen möglich.

2.3.2 Vorhandene praxisnahe Systeme

Es existieren viele praxisnahe Anwendungen, die menschliches Verhalten auf Basis einer Gesichtsanalyse und zumeist auf Basis des FACS interpretieren. In diesem Abschnitt werde ich einen Überblick über diese Systeme geben. Auch ich werde in dieser Arbeit das FACS als Grundlage der Analyse verwenden.

Zunächst gibt es Systeme, die innere Zustände des Menschen anhand einer Interpretation der Mimik extrahieren, jedoch keine weitere Analyse auf Basis der erkannten Zustände durchführen. Kaliouby und Robinson [KR05] beschreiben beispielsweise ein System, das in der Lage ist, mentale Kategorien anhand einer Gesichtsanalyse abzuleiten. Es wurde konsequent weiterentwickelt und kann inzwischen kommerziell erworben werden¹.

Demgegenüber nutzen viele Systeme eine Gesichtsanalyse als Basis, um menschliches Verhalten zu interpretieren und relevante Rückschlüsse im jeweiligen Anwendungsfeld zu ziehen. Exemplarisch beschreibe ich im Folgenden die Anwendung in vier unterschiedlichen Bereichen. Dabei decke ich Gemeinsamkeiten der Szenarien auf und verbinde diese schließlich mit der Anwendung in meiner Arbeit.

¹<https://www.affectiva.com/>

Gesichtsanalyse im Auto

Emotionalität ist gerade im Straßenverkehr weit verbreitet. Auch hier ist ein optimales Nutzungserlebnis in der Mensch-Auto Interaktion wünschenswert. Eyben u. a. [Eyb+10] geben einen Überblick über die Einflüsse von Emotionalität auf Fahrsicherheit und ordnen die Konzepte des *Affective Computing* im Anwendungsgebiet des Autos ein. Auch Grimm u. a. [Gri+07] arbeiten detailliert heraus, dass eine Analyse des emotionalen Zustands des Autofahrers wichtig sei. Gerade im Bereich der Fahrerbeobachtung gibt es bereits eine Vielzahl von Systemen, die auf einer automatischen Analyse des Gesichtes basieren. Dieses ist nicht verwunderlich, da das Gesicht des Fahrers während der gesamten Fahrzeit annähernd am selben Ort und wenn überhaupt nur selten verdeckt ist. Somit sind die Umgebungsbedingungen optimal, um Gesichtsanalysen für einen Erkenntnisgewinn über den Zustand des Fahrers zu verwenden.

Viele Systeme wurden entwickelt, um *Aufmerksamkeit* [Ber+06; Don+11; LRL07], *Müdigkeit* [GS01; LC12; Qio+06; Vur+07], *Frustration* [Mal+11], *Stress* [HP05] oder *mentale Belastung* [PGB14; Yuc+17] im Auto zu erkennen. Dabei ist die Gesichtsanalyse das einzige Merkmal oder nimmt eine wichtige Rolle in einer multimodalen Analyse ein.

Neben der Schätzung von inneren menschlichen Zuständen könne auch zukünftiges Fahrverhalten vorhergesagt werden. So beschreiben Jabon u. a. [Jab+11] ein Verfahren, um anhand einer Gesichtsanalyse *unsicheres* Fahrverhalten in einem Fahrsimulator vorherzusagen. Auch *Abbiegeintentionen* könnten anhand eines multimodalen Verfahren auf Basis einer Gesichtsanalyse erkannt werden [Jai+15].

Gesichtsanalyse im medizinischen Umfeld

Die Analyse von Gesichtern kann auch im der Medizin eingesetzt werden. Ein dauerhafter Einsatz der Systeme in der Praxis konnte bisher jedoch nicht festgestellt werden. Hammal und Cohn [HC18], Ashraf u. a. [Ash+09] und Lucey u. a. [Luc+11] stellen Systeme vor, die in der Lage sind, anhand einer Gesichtsanalyse *Schmerzen* von Versuchspersonen zu erkennen.

Das starker Schmerz auch durch Mimik gezeigt wird, ist durchaus nachvollziehbar. Eine weitere und weitaus komplexere Anwendung im medizinischen Umfeld ist die Erkennung von *Depressionen*. Cohn u. a. [Coh+18] beschreiben ein Verfahren, das Depressionen in

Interviewsituationen anhand einer Gesichtsanalyse in Kombination mit auditiven Merkmalen erkennen kann.

Auch bei den genannten Anwendungen im medizinischen Umfeld ist festzustellen, dass die betrachteten Situationen in einer definierten Umgebung stattfinden. Die Probanden befinden sich jeweils zentral vor einem Kamerasystem und bewegen sich nur wenig.

Gesichtsanalyse in Lernumgebungen

Im Bereich der Lernumgebungen wird die Gesichtsanalyse besonders beim elektronisch unterstützten Lernen betrachtet. Es wird untersucht, ob anhand des Gesichtes feststellbar ist, ob die lernende Person noch konzentriert dem Lernstoff folgt und nicht bereits überfordert oder gelangweilt ist. Gerade durch die starke Verbreitung von internationalen virtuellen Lernplattformen (wie beispielsweise die Plattform *Udacity*²) mit tausenden Lernenden pro Einheit ist eine automatische Analyse des inneren menschlichen Zustandes nicht zuletzt als ein Faktor zur Lernfortschrittskontrolle wünschenswert, da keine direkte Kontrolle durch die Lehrkraft mehr möglich ist.

So werde in Lernumgebungen bereits *Frustration* [Gra+13], *Engagement* [Whi+14] und *Affekt* [BR11] auf Basis einer Gesichtsanalyse erkannt. Auch Systeme zur Erkennung von multiplen mentalen Kategorien wurden bereits beschrieben [DPG07; Cra+04]. Auf der Basis einer Zustandsanalyse können Architekturen entwickelt werden, um Emotionen und innere Zustände in Lernsysteme zu integrieren und somit adäquat auf die lernende Person reagieren zu können [KN16].

Gesichter können in Lernumgebungen als Kanal verwendet werden, um Rückschlüsse auf den Lernerfolg und den inneren Zustand der lernenden Person zu ziehen. Auch in diesem Fall ist die Umgebungssituation klar definiert. Die lernende Person befindet sich vor einem Computer und bewegt sich nur wenig, da sie dem Lernmaterial auf dem Bildschirm folgt. Eine mögliche Störung wird von Whitehill u. a. [Whi+14] beschrieben. Sie stellen fest, dass Probanden oftmals die Hand vor dem Mund halten und dieses die Erkennung der Mimik beeinflusse. Störungen dieser Art können auch im entwickelten Szenario dieser Arbeit auftreten.

Wiederum fällt die klar definierte und bekannte Umgebung auf, in der die Analysesysteme eingesetzt werden. Im Verlauf dieser Arbeit werde ich eine Lernumgebung verwenden,

²<https://www.udacity.com/>

um zu zeigen, dass die von mir entwickelten Verfahren der Prozessanalyse generalisierbar sind und neben der Anwendung in einer industriellen Montageumgebung auf weitere Bereiche übertragen werden können.

Gesichtsanalyse in der Marktforschung

Ein Ziel der Marktforschung ist es herauszufinden, wie Menschen Produkte oder Medieninhalte einschätzen. So kann beispielsweise das Kaufverhalten bestimmter Personengruppen abgeschätzt werden [Tho+17, Kapitel 5.1]. Durch die Verwendung von Gesichtsanalyse ist es möglich eine direkte Rückmeldung zu den gezeigten Inhalten zu bekommen. Eine solche Analyse ist ähnlich zu den in dieser Arbeit vorgeschlagenen Konzepten zur Bewertung von Assistenzsystemen. Durch eine Gesichtsanalyse muss die Versuchsperson auch in der Marktforschung keine Fragebögen ausfüllen, sondern ihre Bewertung kann direkt beim Konsum der Medieninhalte abgeleitet werden.

Joho u. a. [Joh+11] stellen ein System dar, dass in der Lage ist durch die Analyse des Gesichtes persönliche Highlights in Videos zu finden. Dabei seien Bewegungen im oberen Teil des Gesichtes (Augen, Wange) besser für eine Analyse geeignet als Bewegungen im unteren Bereich des Gesichtes (Mund). Das von McDuff u. a. [McD+15] vorgestellte Verfahren sticht im Vergleich zu anderen Systemen besonders heraus. Es sei in der Lage anhand einer Analyse des Gesichtes beim Konsumieren von Werbefilmen, Rückschlüsse auf Gefallen und Kaufabsichten der gezeigten Produkte abzuleiten. Dabei ist besonders die Größe der erhobenen Datenbasis mit 12230 Videoaufnahmen von 1223 Probanden aus vier Ländern beeindruckend. Darüber hinaus ist eine Vorhersage von Filmbewertungen durch ein von Navarathna u. a. [Nav+17] entwickelten System möglich. Dabei wird das Verhalten von Kinobesuchern über die gesamte Filmdauer analysiert, wobei die Gesichtsanalyse in einem multimodalen Setup eine Ausnahmestellung einnimmt.

2.3.3 Diskussion

Es existieren bereits viele praxisnahe Entwicklungen, die eine Analyse des Gesichtes verwenden, um den inneren Zustand des Menschen abzuleiten. Darüber hinaus gibt es bereits Anwendungen, die durch eine menschliche Verhaltensanalyse Inhalte bewerten und Vorhersagen über zukünftiges Verhalten und Meinungen des Menschen geben können.

Über die verschiedenen vorgestellten praxisnahen Anwendungsgebiete hinweg lässt sich feststellen, dass die Szenarien deutliche Ähnlichkeiten aufweisen. Zunächst bewegt sich die zu beobachtende Person kaum. Dieses macht eine Analyse mit einem fest installierten Kamerasystem möglich. Auch in dem von mir entwickelten Montageassistenzsystem bewegt sich die zu beobachtende Person kaum, da sich diese an einem festen Arbeitsplatz befindet. Somit kann eine Gesichtsanalyse ohne nennenswerte Einschränkungen oder Anpassungen angewendet werden.

Außerdem fällt bei der Analyse der praxisnahen Entwicklungen auf, dass der innere Zustand der zu analysierenden Person unerheblich ist. Es ist für die eigentliche Anwendung unwichtig, welche psychologische Kategorie das gezeigte Verhalten ausgelöst hat. Vielmehr ist es wichtig, dass das gezeigte Verhalten der nutzenden Personen mit der zu lösenden Aufgabe bzw. der Mensch-System Interaktion in Verbindung gebracht werden kann. Der innere menschliche Zustand kann demzufolge als *Black Box* gesehen werden und eine Benennung dieses Zustandes ist nicht nötig.

Besonders die Anwendung in der Marktforschung ähnelt den von mir entwickelten Verfahren zur Analyse von Assistenzsystemen und -prozessen. Die von mir vorgestellte Verwendung einer Gesichtsanalyse als Evaluationsverfahren in der nutzerorientierten Entwicklung von prozessbasierten Systemen wurde meines Wissens noch nicht betrachtet.

Die vielen Beispiele einer robusten und erfolgreichen Verwendung der Gesichtsanalyse lassen darauf schließen, dass auch der Einsatz im industriellen Umfeld zur Bewertung von Assistenzsystemen und Prozessschritten gute Ergebnisse hervorbringen wird.

2.4 Zusammenfassung

In diesem Kapitel habe ich zunächst dargestellt, dass besonders im industriellen Umfeld die Nachfrage nach guten und optimalen Assistenzsystemen und -prozessen immer größer wird und in Zukunft mit einer weiteren Steigerung der Nutzung von Assistenzsystemen gerade in Montageumgebungen gerechnet werden muss. Um optimale Prozesse zu entwickeln und die Zufriedenheit bei der Assistenzsystemnutzung zu maximieren, habe ich das Konzept der nutzerorientierten Entwicklung eingeführt und dabei auf Lücken bei den Evaluationsmöglichkeiten hingewiesen. Besonders bei der Quantifizierung des Nutzungserlebnisses und damit der Berücksichtigung der Emotionalität und des inneren

Zustandes der nutzenden Person werden bislang subjektive Einschätzungen der Versuchspersonen verwendet. Ich schlage vor, stattdessen Analyseverfahren zu verwenden, die menschliches Verhalten während der Systeminteraktion direkt analysieren, um die genannten subjektiven Einschätzungen durch eine objektive Bewertung zu ersetzen. Durch die Beschreibung praxisnaher Beispiele motiviere ich besonders die Benutzung der Gesichtsanalyse als aussagekräftigen Merkmalskanal, da es bereits in anderen Bereichen möglich ist, innere menschliche Zustände anhand der Mimik abzuleiten und daraus Rückschlüsse zu ziehen.

Praxisnahe Anwendung der nutzerorientierten Entwicklungskonzepte

„Jedes Denken wird dadurch gefördert, dass es in einem bestimmten Augenblick sich nicht mehr mit Erdachtem abgeben darf, sondern durch die Wirklichkeit hindurch muss.“

– Albert Einstein

Um ein möglichst praxisnahes Szenario zu entwickeln, habe ich eine enge Zusammenarbeit mit dem *Technologie-Netzwerk Intelligente Technische Systeme OstWestfalenLippe (it's OWL)* aufgebaut. In diesem Netzwerk sind etwa 200 Unternehmen, Forschungseinrichtungen und Organisationen vereint. Sie entwickeln „[...] gemeinsam neue Technologien für die Fabrik der Zukunft“ [its18]. In diesem Netzwerk werden *Transferprojekte* gefördert, die sich durch eine enge Zusammenarbeit zwischen Wissenschaft und Wirtschaft auszeichnen, um erlangtes Wissen aus der Wissenschaft vor allem in kleine und mittlere Unternehmen zu transferieren. Ich habe dabei eng mit dem Transferprojekt *Interaktive Mitarbeiterassistenz für die Elektromotorenmontage (IMEMO)* kooperiert und konnte dadurch eine konkrete Umsetzung im Unternehmen *HANNING ELEKTRO-WERKE GmbH & Co. KG (HANNING)* mitgestalten.

Die Firma HANNING (weltweit etwa 1500 Mitarbeitende) ist ein führender Hersteller kundenspezifischer elektrischer Antriebssysteme. Es werden pro Jahr etwa 4000 unterschiedliche Motorvarianten produziert. Dabei variieren die Losgrößen zwischen 5 und etwa 1000. Die hohe Komplexität der Montageaufgabe, die kleinen Losgrößen der Motoren und der unterschiedliche Wissensstand der Belegschaft führen zu Herausforderungen in

der manuellen Endmontage. Dadurch kann es sowohl zu deutlichen Einbußen der Effizienz, als auch zu einem steigenden Risiko von Montagefehlern kommen.

Zu Beginn des Projektes wurden die Montageanleitungen für die Endmontage in Papierform zur Verfügung gestellt. Die arbeitenden Personen wurden durch kein digitales System unterstützt. Diese Situation diente als Startpunkt, um einen nutzerorientierten Entwicklungsprozess eines Assistenzsystems durchzuführen und darauf folgend die in dieser Arbeit entwickelten Verfahren zu evaluieren.

In diesem Kapitel stelle ich die Anwendung der nutzerorientierten Entwicklung am Beispiel der Assistenzsystementwicklung in manuellen Montageumgebungen dar. Durch die enge Zusammenarbeit mit der Firma HANNING hat die Umsetzung eine hohe Relevanz für die Praxis und dient außerdem als Basis einer wissenschaftlichen Evaluationsplattform für die entwickelten Ansätze.

3.1 Projektverlauf

Damit ein Assistenzsystem den Ansprüchen der Belegschaft gerecht wird und diese bei ihren individuellen Herausforderungen unterstützt, wurde der beschriebene nutzerorientierte Systementwicklungsprozess (siehe Abschnitt 2.2, Seite 10) angewendet.

Zunächst wurden Workshops durchgeführt, an denen unterschiedliche Fachabteilungen (u. a. Arbeitsvorbereitung, Montageleitung) und das System zukünftig nutzende Personen teilgenommen haben. Dabei wurden unter Berücksichtigung der Erfahrungswerte, Anforderungen an das System definiert und Verbesserungspotentiale der aktuellen Montagesituation ermittelt. Außerdem wurde gemeinsam ein Prozess der aktuellen Fertigung ausgewählt, der mit dem zu entwickelnden System prototypisch als erstes umgesetzt werden soll. Dieser deckt die nötige Komplexität ab und stellt die Belegschaft bereits jetzt vor große Herausforderungen.

3.1.1 Modellierung des Montageprozesses

Nach dieser Konzepterstellung wurde der identifizierte Prozess modelliert und in einen Prototypen umgesetzt. Die folgende formale Beschreibung wurde als Grundlage einer

klaren Definition des Prozessablaufs, zur Vereinfachung des zu entwickelnden Systems und für einen gemeinsamen Ausgangspunkt der Entwicklung erstellt:

- Annahme 3.1** Die Montage wird als Prozess P definiert, welcher aus n Gesamtschritten besteht.
- Annahme 3.2** Nur ein Arbeiter M wird die Verkablung des Motors durchführen.
- Annahme 3.3** Jeder Montageschritt i von n Gesamtschritten wird einzeln durchgeführt.
- Annahme 3.4** Die Montageschritte werden in einer zu definierenden Reihenfolge durchgeführt.
- Annahme 3.5** Die Montageschritte werden an drei unterschiedlichen Stationen durchgeführt.
- Annahme 3.6** Es gibt einen definierten Startzustand S .
- Annahme 3.7** Es gibt einen definierten Endzustand Z .
- Annahme 3.8** Alle Teile werden während der Montage fest miteinander verbunden.

Dem folgend konnten in weiteren durchgeführten Workshops Anforderungen erarbeitet werden, die das zu entwickelnde Assistenzsystem erfüllen muss:

- Anforderung 3.1** Die Modellierung von Bauteilgruppen soll generisch und für andere Szenarien wiederverwendbar sein.
- Anforderung 3.2** Verschiedene arbeitende Personen sollen je nach Erfahrung einen unterschiedlichen Grad der Unterstützung erhalten.
- Anforderung 3.3** Die Montage soll zur Qualitätssicherung automatisch dokumentiert werden.

Durch Annahme 3.1 kann auf standardisierte Prozessmodellierungsmethoden zurückgegriffen werden. Dabei wird, wie auch schon von Rütter u. a. [Rüt+13] beschrieben, der *Business Process Model and Notation 2.0 (BPMN 2.0) Standard* [Obj09; CT12] verwendet. Damit werden Prozesse in einer grafischen Spezifikationssprache beschrieben. Der BPMN 2.0 Standard kommt bislang vor allem in der Modellierung ganzheitlicher Geschäftsprozesse zum Einsatz. Er deckt aber auch die Anforderungen für die in diesem Fall benötigten Montageprozesse ab. Zur konkreten Umsetzung wurde die frei verfügbare *Camunda Workflow Engine*¹ verwendet. Im Folgenden werde ich den BPMN 2.0 Standard näher erläutern.

¹<https://camunda.com/>



(a) Aufgaben, die durch die nutzende Person erfüllt werden



(b) Aufgaben, die durch das System erfüllt werden

Abbildung 3.1.: Aufgabentypen nach BPMN 2.0 Standard:

Es werden die nach BPMN 2.0 Standard modellierbaren Aufgabentypen dargestellt.

Die Grundlage des BPMN 2.0 Standards sind *Aktivitäten*. Sie repräsentieren abgeschlossene Aufgaben, die entweder von einer Person (siehe Abbildung 3.1a) oder dem System (siehe Abbildung 3.1b) ausgeführt werden. Nutzeraufgaben muss die bedienende Person erfüllen und nach Erledigung bestätigen (z. B. „Anziehen einer Schraube“). Systemaufgaben werden vom System ausgeführt und nach deren Ausführung wird automatisch in den nächsten Prozessschritt gewechselt (z. B. „Erstellung einer Datei im Dateisystem“).



Abbildung 3.2.: Beispielhafter sequentieller Prozess:

Durch die Ausführung von Montageschritten (Schritt 1 – Schritt n) wird von einem Startzustand *Start S* (Annahme 3.6) ausgehend ein Endzustand *Ziel Z* (Annahme 3.7) erreicht.

Annahme 3.2, Annahme 3.3 und Annahme 3.4 lassen eine deutliche Vereinfachung der zu erstellenden Prozessmodelle zu. Der Prozess kann durch eine sequentielle Abfolge von Montageschritten modelliert werden (siehe Abbildung 3.2) und es müssen keine parallelen Prozessschritte berücksichtigt werden. In Abbildung 3.2 ist außerdem der in Annahme 3.6 geforderte Startzustand und in Annahme 3.7 geforderte Endzustand zu sehen. Diese garantieren einen definierten Start und ein definiertes Ende des Prozesses. Auch Annahme 3.8 vereinfacht die Modellierung, da durch die feste Verbindung der Teile keine bidirektionalen Kanten im definierten Prozess auftreten können. Bildlich gesprochen muss somit kein „Schritt zurück“ modelliert werden. Eine bidirektionale Kantenmodellierung ist des Weiteren im BPMN 2.0 Standard nicht vorgesehen. Der Prozessfluss ist immer direktional. Eine Berücksichtigung könnte höchstens durch explizite Modellierung realisiert werden. Dadurch würde zu jedem Montageschritt eine Entscheidungsfunktion hinzugefügt werden, um somit eine zweite ausgehende Kante zu dem vorherigen Schritt zu erstellen. Das entstehende Prozessmodell würde dadurch jedoch unnötig komplex.

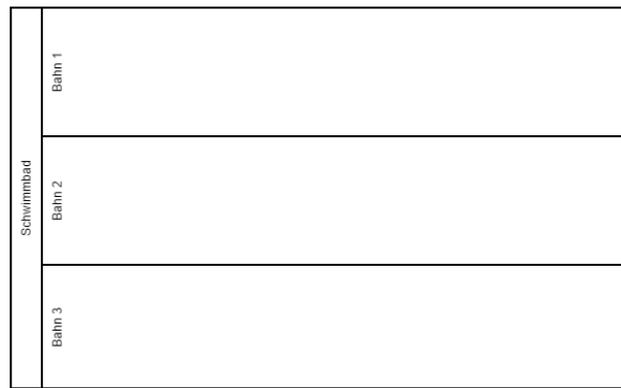


Abbildung 3.3.: Schwimmbadmodellierung nach BPMN 2.0 Standard:

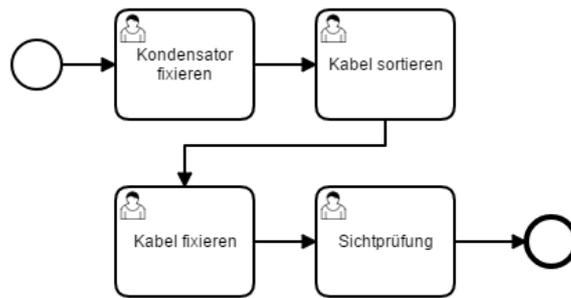
Es wird ein Schwimmbad nach dem BPMN 2.0 Standard mit drei enthaltenen Bahnen dargestellt.

Um Annahme 3.5 zu genügen, wurden die drei Montagestationen explizit in die Prozessmodellierung integriert. Der BPMN 2.0 Standard enthält *Schwimmbäder* (engl. *Pools*) mit enthaltenen *Bahnen* (engl. *Lanes*) (siehe Abbildung 3.3). Im eigentlichen Sinne repräsentiert das Schwimmbad eine beteiligte Prozessinstanz wie etwa ein Unternehmen. Werden teile des Prozesses in unterschiedlichen Unternehmen durchgeführt, kann es auch zum Vorkommen von verschiedenen Schwimmbädern kommen, die jeweils ein unabhängiges Unternehmen repräsentieren. Ein Schwimmbad kann dabei in eine beliebige Anzahl Bahnen unterteilt werden. Diese repräsentieren unterschiedliche *Teilnehmer* (engl. *Participants*) am Prozess und können beispielsweise einzelne Fachabteilung eines Unternehmens sein. Dieses Konzept wurde für das Anwendungsszenario in dieser Arbeit adaptiert. Hier repräsentiert das Schwimmbad einen Arbeitsplatz, an dem eine Person M arbeitet. Dieser Arbeitsplatz besteht aus drei Bahnen, die jeweils einen einzelnen Arbeitsbereich darstellen.

Anforderung 3.1 kann auch durch eine Verwendung von Elementen des BPMN 2.0 Standards erfüllt werden. Es besteht die Möglichkeit *Subprozesse* zu definieren (siehe Abbildung 3.4). Diese sind in sich abgeschlossen und bestehen aus Teilaufgaben und eigener Start- und Endzustände. Einerseits können sie expandiert dargestellt werden (siehe Abbildung 3.4b). Diese Darstellung unterscheidet sich nicht von der oben beschriebenen Darstellung von Prozessen. Andererseits können Subprozesse kollabiert modelliert werden (siehe Abbildung 3.4a). Diese Art der Modellierung ermöglicht es – wie gefordert – wiederverwendbare und generische Modellgruppen zu modellieren. Es sind auch hierarchische Strukturierungen möglich, die Montagegruppen aus einem Zusammenspiel bereits existierender Untermontagegruppen modellieren. Subprozesse können darüber hinaus mit



(a) Kollabierte Darstellung eines Subprozesses



(b) Expandierte Darstellung eines Subprozesses

Abbildung 3.4.: Subprozessmodellierung nach BPMN 2.0 Standard:

Es wird die kollabierte und die expandierte Modellierung eines Subprozesses nach dem BPMN 2.0 Standard dargestellt.

Variablendefinitionen parametrisiert werden. Dieses erhöht die Generalisierungsfähigkeit weiter, wurde aber im Anwendungsszenario nicht umgesetzt und bleibt für zukünftige Entwicklungen zu berücksichtigen.

Anforderung 3.2 erfordert eine Erweiterung des BPMN 2.0 Standards, indem im entwickelten System verschiedene Arten der bereits beschriebenen Nutzeraufgaben verwendet werden können. Diese Erweiterung wurde umgesetzt, indem zum einen *sicherheitsrelevante Nutzeraufgaben*, zum anderen *normale Nutzeraufgaben* definiert werden können. Alle sicherheitsrelevanten Aufgaben müssen von jeder Person – egal welche Erfahrung diese hat – durchgeführt und explizit bestätigt werden. Die normalen Aufgaben werden bei erfahrenen Mitarbeitern automatisch zusammengefasst und müssen nicht mehr jeweils explizit bestätigt werden. Somit wird ein erfahrener Nutzer nicht vom System aufgehalten, da er einfache und bereits bekannte Aufgabenschritte nicht wiederholt bestätigen muss. Exemplarisch wurde im entwickelten System daher die beiden Nutzerrollen *viel Erfahrung* und *wenig Erfahrung* modelliert und implementiert. Diese vereinfachte Berücksichtigung des Erfahrungslevels muss in Zukunft deutlich weiterentwickelt werden. Für eine Erprobung und Verwendung in der realen Montage reicht die genannte Modellierung aber völlig aus.

Zur automatischen Dokumentation der Montage (Anforderung 3.3) wurden Prinzipien entwickelt, die einen Bericht als PDF Dokument oder als XML Datei exportieren. Dieser Export ist frei konfigurierbar. Es können beispielsweise gemachte Prüfbilder (Zwischenaufnahmen vom Montagezustand) oder benötigte Zeiten für die einzelnen Teilschritte exportiert und gesichert werden.

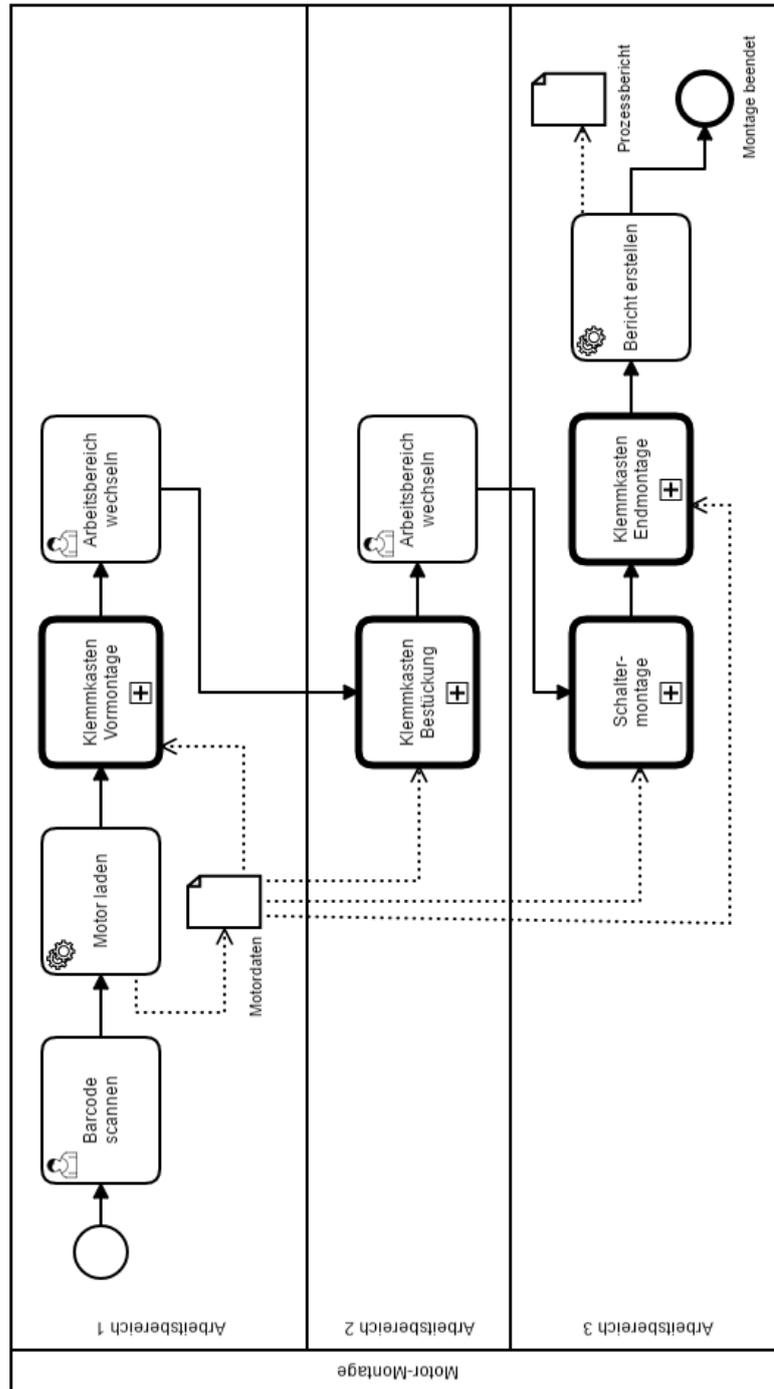


Abbildung 3.5.: Prozessablauf des modellierten Anwendungsszenarios:

Es werden drei Arbeitsbereiche, unterschiedliche System- und Nutzeraufgaben und kollabierte Subprozesse verwendet. Außerdem sind die benötigten Motordaten, der erstellte Prozessbericht, der Startzustand und der Endzustand visualisiert. Sämtliche Elemente verwenden den BPMN 2.0 Standard.

Name	Anzahl
Systemaufgaben	6
Nutzeraufgaben	116
Subprozesse	22
Hierarchische Tiefe	4
Medienelemente	236

Tabelle 3.1.: Prozessdaten des Anwendungsszenarios:

Es wird eine Zusammenfassung der erstellten Aufgabenschritte dargestellt.

Der gesamte Prozess, der auf Basis der beschriebenen Elementen des BPMN 2.0 Standards modelliert wurde, ist in Abbildung 3.5 zu sehen. In Tabelle 3.1 ist ein Überblick über die dabei implementierten Prozessdaten dargestellt. Der modellierte Montageprozess besteht aus 6 unterschiedlichen Systemaufgaben und 116 Nutzeraufgaben mit 236 annotierten Medienelementen. Die Aufgaben sind in 22 Subprozesse mit einer maximalen hierarchischen Tiefe von 4 gruppiert.

3.1.2 Entwicklung der Nutzeroberfläche

Neben der Prozessmodellierung durchlief auch die Nutzeroberfläche einige Iterationen der nutzerorientierten Entwicklung. Durch ständige Kommunikation und Evaluation mit der Belegschaft der Firma HANNING wurde sie mehrfach angepasst und weitentwickelt. Ein anfänglicher Prototyp ist in Abbildung 3.6 zu sehen. Demgegenüber ist in Abbildung 3.7 die finale Version dargestellt, die stark an die Bedürfnisse der Personen, die das System nutzen, angepasst wurde.

Diese Entwicklung wäre ohne einen starken Einbezug der nutzenden Personen nicht möglich gewesen. So wurden die Navigationselemente auch farblich deutlich voneinander getrennt. Außerdem wurden wichtige und häufig zu verwendende Funktionalitäten hervorgehoben und vergrößert (siehe „Aufgabe erledigt“ Knopf zum Bestätigen der aktuellen Aufgabe).

Die Nutzer sind des Weiteren in der Lage, Bilder zu vergrößern und in gewünschte Bereiche zu zoomen (siehe Abbildung 3.8). Auch die klar strukturierte Darstellung des aktuellen Prozesses und der Position in diesem Prozess (siehe Abbildung 3.7, linke Seite) wurde auf Wunsch der Endanwender hinzugefügt.

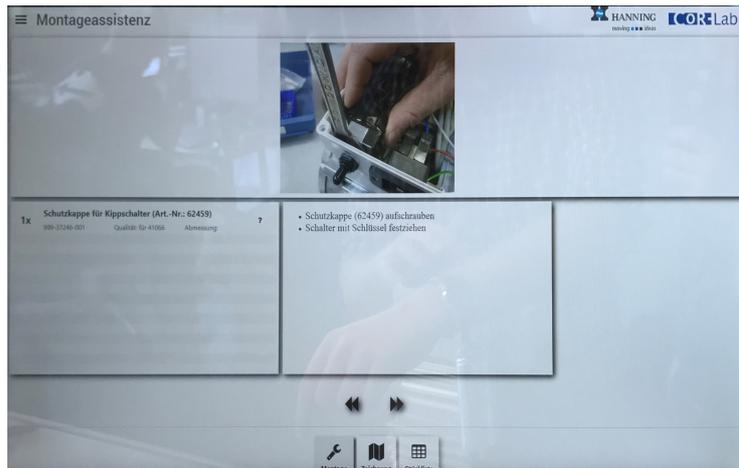


Abbildung 3.6.: Nutzeroberfläche des ersten Prototypen:

Die Nutzeroberfläche besteht aus einer Liste der relevanten Montageteile und aus einer Darstellung der aktuellen Aufgabe als Bild und als eine stichpunktbasierte Beschreibung. Neben der Bestätigung des aktuellen Schrittes und dadurch bedingt dem Weiterschalten zum nächsten Prozessschritt kann die Dokumentation des vorherigen Schrittes aufgerufen werden. Außerdem können allgemeine Darstellungen wie Übersichtszeichnungen und -stücklisten der Baugruppe angezeigt werden.

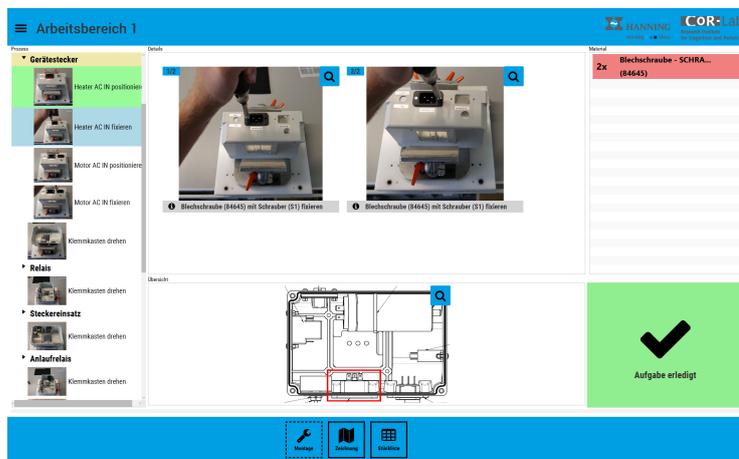


Abbildung 3.7.: Nutzeroberfläche des finalen Systems:

Im Vergleich zu Abbildung 3.6 sind die zu verwendenden Montageteile farblich rot hervorgehoben und der gesamte Prozessablauf wird am linken Rand der Oberfläche dargestellt. Hier kann auch die Dokumentation beliebiger Schritte aufgerufen werden. Darüber hinaus kann eine Zeichnung zu jedem einzelnen Schritt hinzugefügt werden. Die Bestätigung des Schrittes wurde durch den grünen Knopf deutlich sichtbarer gemacht und die textuelle Beschreibung der Aufgabe wurde standardisiert.

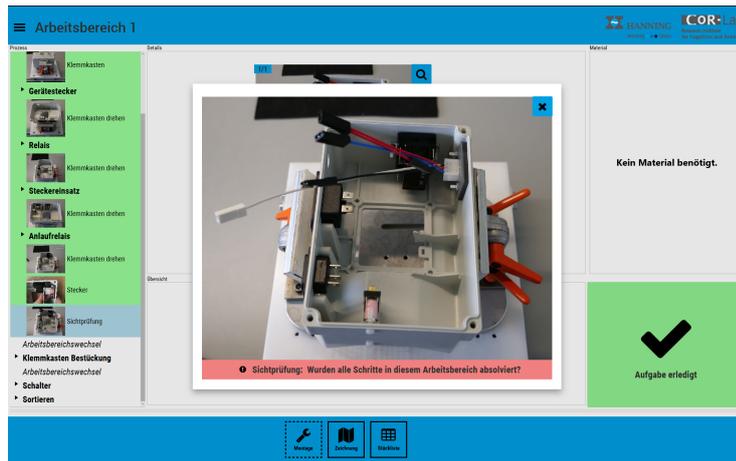


Abbildung 3.8.: Bildvergrößerung in der Nutzeroberfläche:

Sämtliche Abbildungen, Diagramme und Zeichnungen können durch Touch-Interaktion vergrößert und gezoomt werden.

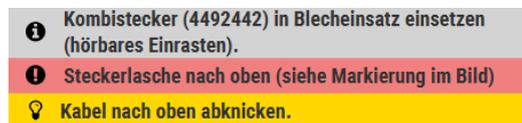


Abbildung 3.9.: Beschreibungsarten eines Montageschrittes:

Ein Montageschritt wird standardisiert beschrieben. Dabei werden drei Arten der Unterscheidung durch unterschiedliche Farben und verschiedene Symbole verdeutlicht. Informationen sind grau, Tipps sind gelb und Warnungen sind rot dargestellt.

Weiterhin wurde eine standardisierte Erstellung von Montageschritten durch die Belegschaft vorgeschlagen und darauf aufbauend implementiert. Pro Aufgabenschritt sind bis zu drei Arten der Beschreibung möglich (*Information*, *Tipp* und *Warnung*). Dieser Sachverhalt ist beispielhaft in Abbildung 3.9 dargestellt. Informationen werden durch einen grauen, Tipps durch einen gelben und Warnungen durch einen roten Hintergrund visualisiert. Zu jedem Montageschritt kann der Modellierer darüber hinaus eine beliebige Anzahl von Medienelementen wie etwa Bilder hinzufügen. Dabei können diese zusätzlich intuitiv mit Annotationen versehen werden.

Eine weitere gewünschte und umgesetzte Funktionalität ist der schnelle Wechsel in eine Dokumentation, um sich vergangene oder zukünftige Montageschritte anzuschauen. Am eigentlichen Prozessfortschritt und der aktuellen Position im Prozess werden durch das System keine Änderungen vorgenommen (siehe ockerfarbene Markierung des aktuellen Schrittes im Prozessablauf in Abbildung 3.10, linke Seite). Falls die nutzende Person die Dokumentation verlassen will, wird durch das Betätigen des gelben Knopfes (Abbildung 3.10,

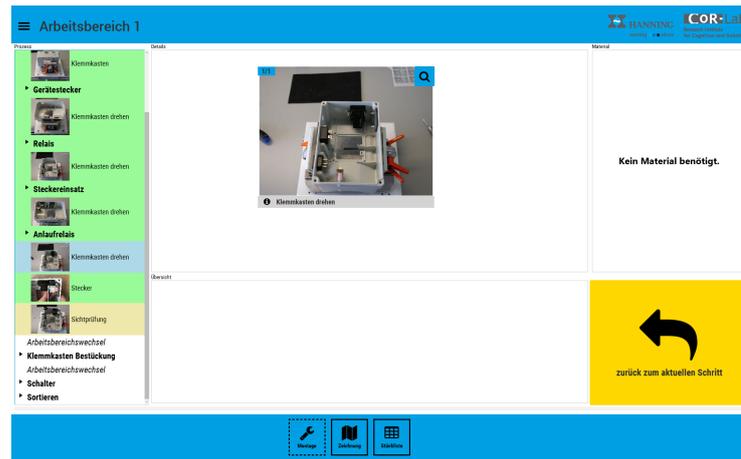


Abbildung 3.10.: Zugriff auf die Dokumentation in der Nutzeroberfläche:

Die Dokumentation ist nahtlos in der Oberfläche integriert. Wird diese aufgerufen ist der Prozessschritt im Prozessablauf auf der linken Seite blau markiert. In Ocker ist der aktuelle Prozessschritt zu sehen. Zum Verlassen der Dokumentation muss durch den gelben Knopf zum aktuellen Schritt im Prozess zurückgekehrt werden, um mit diesem fortzufahren.

rechte Seite) zum ursprünglichen und noch immer aktuellen Prozessschritt zurückgekehrt und von diesem Schritt ausgehend der Prozess fortgesetzt.

3.1.3 Aufbau des finalen Assistenzsystems

Auch der Hardwareaufbau wurde mit den nutzerorientierten Entwicklungsprinzipien in verschiedenen Workshops durch mehrere Iterationen von Grund auf konzipiert, angepasst und optimiert. Das Ergebnis ist in Abbildung 3.11 zu sehen. Zu Beginn des Projektes wurde die Entscheidung getroffen, für jeden Arbeitsbereich einen stationären und berührungsempfindlichen Bildschirm zur Steuerung des Systems zu verwenden. Außerdem wurde ein drehbarer Montageschlitten konstruiert und in den Arbeitsplatz integriert, um das Werkstück ohne großen Kraftaufwand zwischen den drei Arbeitsbereichen bewegen zu können. In Abbildung 3.11 ist ein zu montierender Motor auf dem Montageschlitten platziert. Darüber hinaus sind am Arbeitsplatz vier druckluftbetriebene Schrauber unterschiedlicher Größe und Form zur Befestigung von Montageteilen vorhanden. Ein Barcode-Scanner wird zur Identifizierung des aktuellen Montageprozesses anhand eines Strichcodes auf einem gedruckten Auftragsbogen verwendet. Die jeweiligen Montageteile sind den drei Arbeitsbereichen zugeordnet und in unterschiedlichen, mit Artikelnummern beschrifteten

Kästen gelagert. Die Bewegung des Montageschlittens zwischen den Arbeitsbereichen wird durch drei in die Tischplatte integrierte Distanzsensoren automatisch erkannt und der Prozesssteuerung zur Verfügung gestellt.

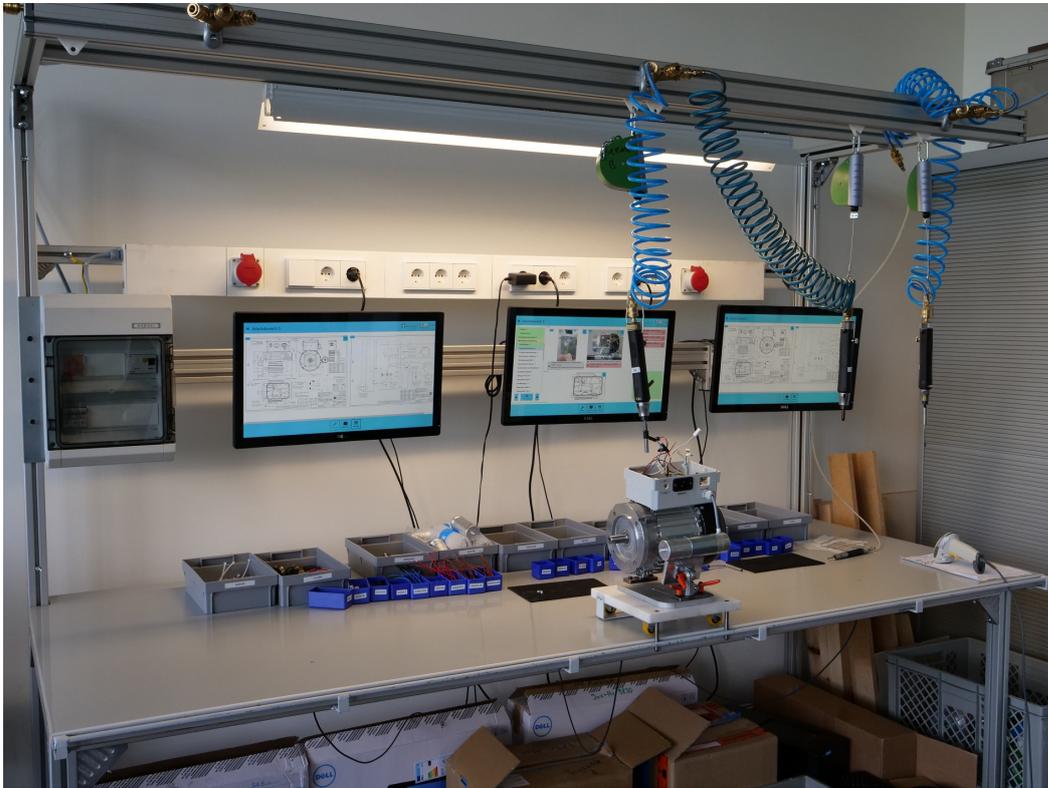


Abbildung 3.11.: Finale Version des entwickelten Arbeitsplatzes:

Dieser Arbeitsplatz wurde mit nutzerorientierten Systementwicklungsprinzipien erstellt. Auf dem Montageschlitten ist der durch die nutzende Person jeweils zu montierende Motor zu sehen.

3.2 Evaluation

Das entwickelte Assistenzsystem wurde abschließend durch zwei Facharbeiter aus der Montage der Firma HANNING evaluiert. Sie wurden nach der Montage eines Motors mithilfe des entwickelten Assistenzsystems befragt und haben des Weiteren den SUS Fragebogen (siehe Kapitel A.1.3, Seite 141) ausgefüllt. Auf einer Basis von zwei Personen kann aus einer SUS Auswertung allenfalls eine sehr grobe Abschätzung der Nutzbarkeit des Systems abgeleitet werden. Dennoch ist es interessant, wie die erfahrenen Facharbeiter die Nutzbarkeit des Systems abschließend bewerten. Das entwickelte System erreicht unter

Berücksichtigung dieser beiden Personen einen SUS von 87.5 und ist somit nach Brooke [Bro13] im *exzellenten* Bereich (Schulnote A) einzuordnen. Der hohe SUS lässt darauf schließen, dass das Assistenzsystem im direkten Vergleich zu der papierbasierten Montage sehr gut abschneidet und die beiden Facharbeiter mit der Nutzbarkeit sehr zufrieden sind.

Auch bei der in Kapitel 5 (siehe Seite 61) beschriebenen Datenerhebung habe ich den SUS berechnet. Dabei konnte eine größere Stichprobe betrachtet werden. Außerdem wurde zusätzlich der RTLX erhoben. Die Ergebnisse sind in Abschnitt 5.3 (siehe Seite 73) beschrieben.

In den zusätzlich durchgeführten Interviews sind durchweg positive Anmerkungen zum System gemacht worden. Es wurde besonders die sehr gute Nutzbarkeit und die extrem einfache und direkt mögliche Übertragbarkeit in die reale Montage der Firma HANNING genannt. Außerdem wurden der Montageschlitten, die einfache Steuerung durch Berührung der Monitore und die übersichtliche Oberfläche gelobt. Die Möglichkeit des Nachschauens von einzelnen Aufgabenschritten in der Dokumentation ist zudem positiv aufgefallen.

Es wurden aber auch hier Verbesserungsvorschläge geäußert. Diese könnten in weiteren Iterationen der nutzerorientierten Entwicklung in Zukunft umgesetzt werden. Besonders ein direktes Anzeigen und somit Transparenz der bereits verbrauchten Zeit pro Montageschritt wurde gefordert. Auch wurde eine mögliche Anpassung der Tischhöhe an verschiedene Körpergrößen genannt.

Die Werkstücke, die die Facharbeiter mit dem Assistenzsystem produziert haben, sind in den normalen weiteren Ablauf der Produktion überführt worden. Sie wurden also von dem Unternehmen geprüft und weiterverarbeitet. Daran lässt sich der Projekterfolg belegen und das entwickelte System kann seinen Zweck erfüllen.

3.3 Zusammenfassung

In diesem Kapitel habe ich den durchgeführten nutzerorientierten Entwicklungsprozess am Beispiel der Assistenzsystementwicklung zum Einsatz in der manuellen Montage vorgestellt. Dabei habe ich eng mit der Firma HANNING zusammengearbeitet, um die Praxistauglichkeit des entwickelten Systems sicherstellen.

In enger Zusammenarbeit mit den Fachabteilungen und den betreffenden Facharbeitenden wurden Annahmen und Anforderungen an das zu entwickelnde System definiert. Der basierende Montageprozess wurde mithilfe einer grafischen Spezifikationssprache (BPMN 2.0 Standard) umgesetzt. Dabei wurden möglichst viele Konzepte des Standards verwendet, dieser wurde aber auch beispielsweise durch eine weitere Differenzierung der Nutzeraufgaben angepasst und erweitert.

Unter Anwendung der nutzerorientierten Entwicklungskonzepte wurde in mehreren Iterationen ein voll funktionsfähiges Assistenzsystem entwickelt. Dabei wurde sowohl der Prozess, der Hardwareaufbau als auch die Nutzeroberfläche stetig weiterentwickelt und angepasst.

Das finale System wurde abschließend von zwei Facharbeitern der Firma HANNING evaluiert. Die Auswertung der Evaluation deutet auf eine sehr gute Nutzbarkeit und Praxistauglichkeit hin.

Merkmalsdefinition zur Verhaltensanalyse in Montageumgebungen

“ *Our ability to imagine complex applications will
always exceed our ability to develop them.*

— Grady Booch

Im vorherigen Kapitel habe ich die Durchführung einer nutzerorientierten Entwicklung eines Assistenzsystems in einer Montageumgebung beschrieben. Ich konnte zeigen, dass das entstandene System nach den Entwicklungszyklen sehr gute Bewertung von erfahrenen Facharbeiten erhält.

Zur besseren Beschreibung der in diesem Kapitel entwickelten Verarbeitungskette zur Merkmalsgenerierung wird der Begriff des Kopfausdruckes verwendet. Ich definiere diesen Begriff wie folgt:

Kopfausdruck Ein Kopfausdruck beschreibt die Kombination der Mimik, der Kopfausrichtung, der Blickrichtung und der Kopfposition.

In diesem Kapitel stelle ich zunächst die verwendeten Merkmale auf Basis einer Analyse des Kopfausdruckes dar. Daraufaufgehend beschreibe ich Merkmale basierend auf weiteren Verhaltensweisen, die der Mensch in Montageumgebungen zeigt. Somit habe ich die Möglichkeit unterschiedlichste Merkmale miteinander kombinieren und vergleichen zu können, um eine Aussage über die am besten zu verwendenden Merkmale für das beschriebene Szenario treffen zu können.

4.1 Merkmalsgruppen auf Basis von Kopfausdrücken

Um den Kopfausdruck verwenden zu können, müssen die folgenden Anforderung erfüllt werden:

Anforderung 4.1 Ein Kamerasystem muss so platziert sein, dass der Kopf der zu beobachtenden Person so häufig und so zentral wie möglich im Kamerabild zu sehen ist.

Anforderung 4.2 Der Kopfausdruck muss robust und valide extrahiert werden können.

Ich habe dazu ein Bildverarbeitungssystem entwickelt, das auf Basis von visuellen Merkmalen des menschlichen Gesichtes den Kopfausdruck extrahieren kann. Ein Überblick des Systems ist in Abbildung 4.1 zu sehen. Die einzelnen Schritte werde ich im folgenden beschreiben. Für eine bessere Übersicht und Referenzierung, sind die einzelnen Schritte in Abbildung 4.1 durch Nummern markiert.

Zunächst wird der Kopfausdruck anhand von AUs, Blickrichtung, Kopfausrichtung und Kopfposition robust geschätzt. Die Kopfausrichtung und -position dienen als Basis einer Aufmerksamkeitserkennung. Abschließend erfolgt eine zeitliche Einbettung, um unterschiedlich lange Zeiträume miteinander vergleichbar zu machen.

4.1.1 Kopfausdruckerkennerung

Die Kopfausdruckerkennerung ist in Abbildung 4.1 mit Markierung 1 gekennzeichnet. Um robuste Merkmale extrahieren zu können (Anforderung 4.2), besteht die Kopfausdruckerkennerung aus einer Vorverarbeitung (Abbildung 4.1, Markierung 1.1) gefolgt von einer Filterung (Abbildung 4.1, Markierung 1.2).

Vorverarbeitung (Abbildung 4.1, Markierung 1.1)

Die Vorverarbeitung verwendet die von Baltrusaitis u. a. [BRM16] vorgestellte Gesichts- und Kopfanalyse¹. Eine Visualisierung sämtlicher erkannter Merkmale der Vorverarbeitung

¹<https://github.com/TadasBaltrusaitis/OpenFace>

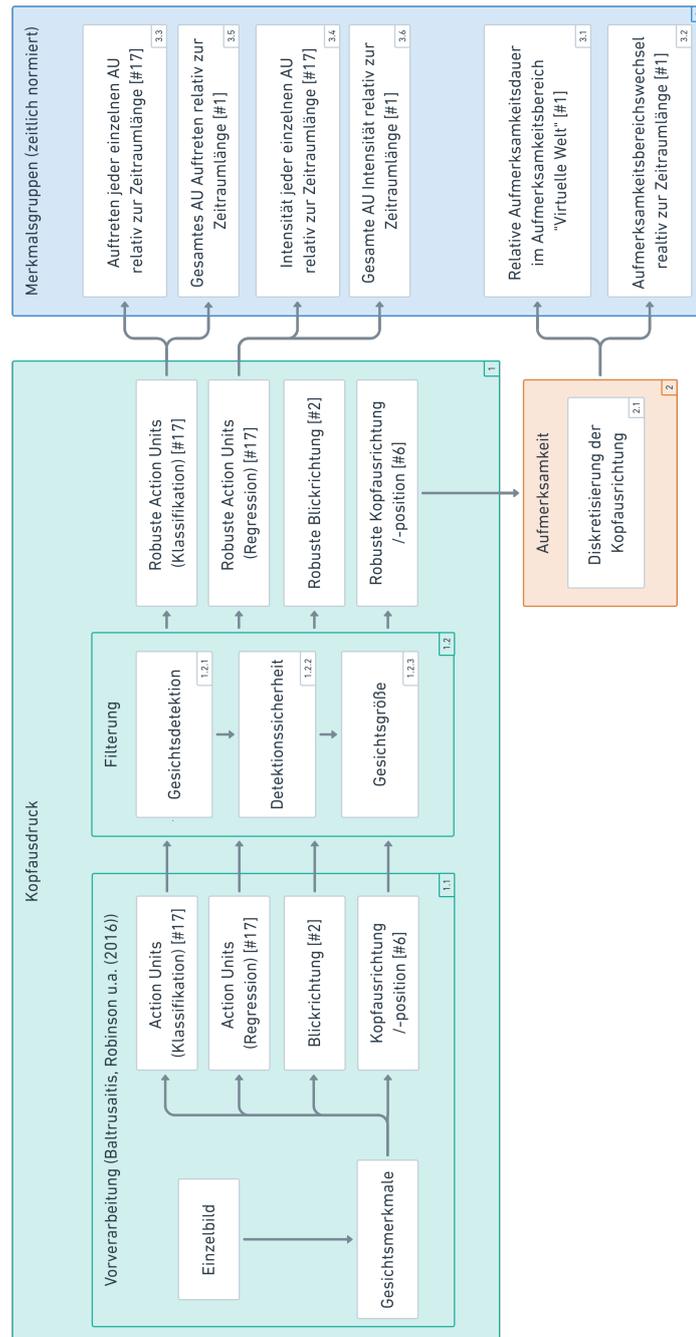


Abbildung 4.1.: Kopfausdruckerkennung und Merkmalsgruppenextraktion:

Der Kopfausdruck wird durch eine Vorverarbeitung [BRM16] und einer mehrstufigen Filterung extrahiert. Darauf aufbauend wird die Aufmerksamkeit erkannt und multimodale zeitlich normierte Merkmalsgruppen ausgegeben. Die Anzahl der Merkmale in den einzelnen Gruppen ist in eckigen Klammern angegeben. Markierungen in den jeweils unteren rechten Ecken der Kästen dienen zur Referenzierung im Text.

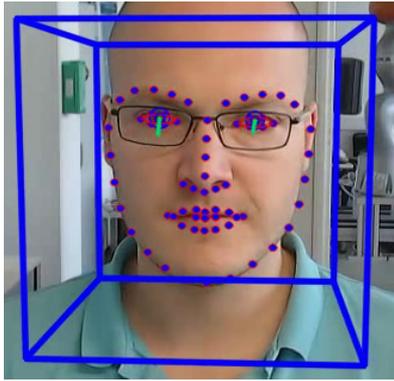


Abbildung 4.2.: Ergebnis der Vorverarbeitung:

Die Kopfausrichtung und Position werden durch einen blauen Quader, die Gesichtspunkte durch blau-rote Punkte, die Augen durch eine rote Umrandung, die Pupillen durch einen blauen Doppelkreis und die Blickrichtung durch grüne Linien verdeutlicht [BRM16].

ist in Abbildung 4.2 zu sehen. Die Kopfausrichtung und die Position werden durch einen blauen Quader, die Gesichtspunkte durch blau-rote Punkte, die Augen durch eine rote Umrandung und die Pupillen durch einen blauen Doppelkreis deutlich gemacht. Außerdem ist die Blickrichtung durch grüne Linien, die am Auge beginnen, dargestellt.

In Abbildung 4.1 ist eine grobe Übersicht der Vorverarbeitung zu sehen. Abbildung 4.3 stellt diese noch einmal detailliert dar. Auch hierbei behalte ich die Methode der Referenzierung einzelner Schritte durch Nummermarkierungen bei und werde im Folgenden auf die einzelnen Verarbeitungsschritte eingehen.

Die Analyse basiert auf einer Betrachtung von Einzelbildern. Für eine Analyse von Videos und Kameraaufnahmen werden diese als eine Abfolge von Einzelbildern betrachtet und jedes einzelne Bild individuell analysiert.

Erkennung von Gesichtern:

(Abbildung 4.3, Markierung 1.1.1)

Der erste Schritt ist die Erkennung eines Gesichtes und die Abschätzung der Gesichtspolition im Bild. Der verwendete Erkennungsalgorithmus basiert auf einer *Max-Margin Objekt Detektion (MMOD)* [Kin15] und verwendet *Histogram of Oriented Gradients (HOG) Merkmale* [DT05]. Diese Merkmale sind in der Lage, Textur zu beschreiben und vergleichbar zu machen. Beispielhaft auf ein Gesicht angewendet sind die extrahierten HOG Merkmale in Abbildung 4.4a visualisiert.

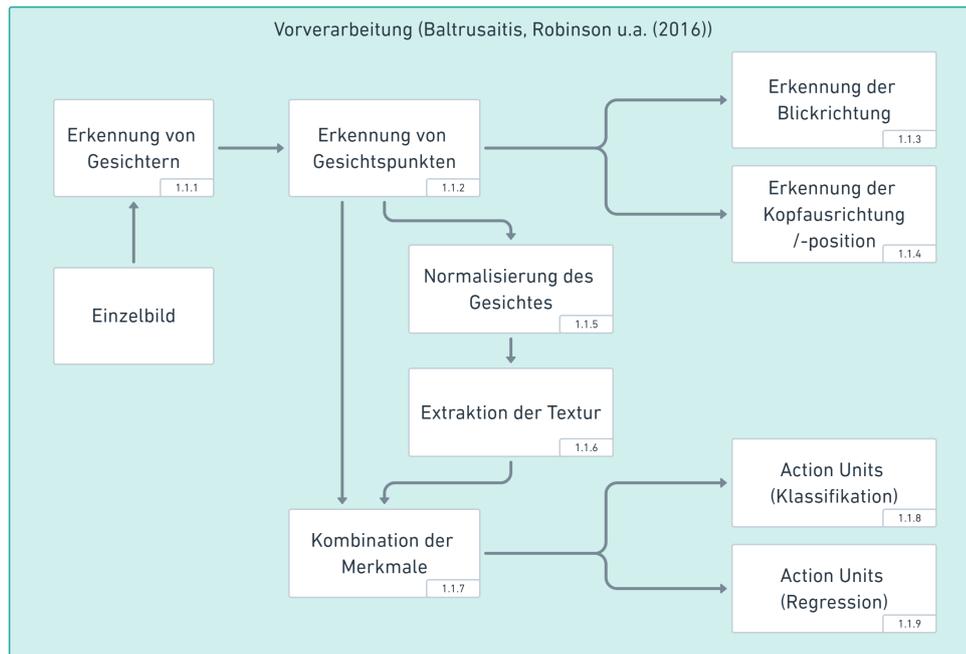


Abbildung 4.3.: Vorverarbeitung der Kopfausdruckererkennung:

Das System erkennt Blickrichtung, Kopfausrichtung, Kopfposition und Action Units (Klassifikation und Regression) [BRM16]. Markierungen in den jeweils unteren rechten Ecken der Kästen dienen zur Referenzierung im Text.

Erkennung von Gesichtspunkten:

(Abbildung 4.3, Markierung 1.1.2)

Das Verfahren zur Erkennung von markanten Punkten in menschlichen Gesichtern ist detailliert von Baltrusaitis u. a. [BRM13] beschrieben. Sie verwenden dazu *Constrained Local Models (CLMs)* [CC06]. Es wird ein globales *Point Distribution Model (PDM)* [CT92] mit lokalen Detektoren verbunden. Für jeden Punkt des PDMs wird ein Experte trainiert, der eine Schätzung der Ähnlichkeit einer lokalen Region zur gelernten Region abgeben kann. Darauf aufbauend kann das PDM optimiert werden, um somit eine Positionsschätzung für jeden einzelnen Punkt unter Berücksichtigung eines globalen Modells abgeben zu können. Ein PDM ist eine formale Beschreibung, wie sich eine definierte Anzahl von Punkten zueinander verhalten. In Abbildung 4.4b ist ein beispielhaftes Modell mit $n = 68$ Punkten visualisiert.

Mathematisch wird ein PDM wie folgt definiert:

$$\mathbf{x}_i = s \cdot \mathbf{R} \cdot (\bar{\mathbf{x}}_i + \Phi_i \mathbf{q}) + \mathbf{t} \quad (4.1)$$

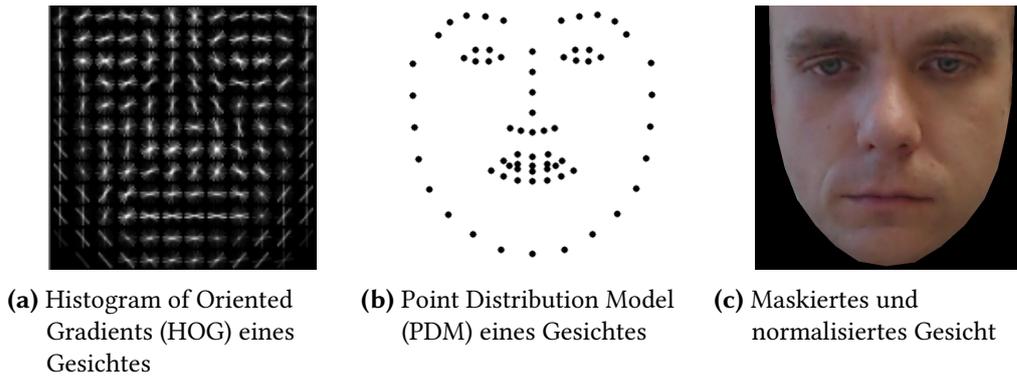


Abbildung 4.4.: Zwischenschritte der Kopfausdruckerennung:

Es werden einzelne Zwischenschritte (Texturmerkmale (HOG), globales Gesichtsmodell (PDM), maskiertes und normalisiertes Gesicht) der Kopfausdruckerennung [BRM16] dargestellt.

Es wird durch den variablen Parameter

$$\mathbf{p} = [s, \mathbf{R}, \mathbf{q}, \mathbf{t}] \quad (4.2)$$

bestimmt. Dadurch lässt sich ein Referenzmodell durch den Skalar s skalieren, durch die Matrix \mathbf{R} rotieren, durch den Vektor \mathbf{t} verschieben und durch den Vektor \mathbf{q} nicht-euklidisch transformieren.

In einer Trainingsphase können nun anhand annotierter Beispiele der Vektor $\bar{\mathbf{x}}_i$ und die Matrix Φ_i gelernt werden. Da ein dreidimensionales Referenzmodell verwendet wird, muss dieses zusätzlich auf zweidimensionale Bildkoordinaten übertragen werden. Hierbei wird von Baltrusaitis u. a. [BRM16] ein Kameramodell auf Grundlage der *skalierten Orthogonalprojektion* verwendet (engl. *weak perspective projection*).

Zur Optimierung des CLMs auf ein zu analysierendes Bild wird das *Regularised Landmark Mean Shift (RLMS)* Verfahren [SLC11] verwendet:

$$\mathbf{p}^* = \arg \min_{\mathbf{p}} \left[\mathcal{R}(\mathbf{p}) + \sum_{i=1}^n \mathcal{D}_i(\mathbf{x}_i; \mathcal{I}) \right] \quad (4.3)$$

Dabei wird die Fehlerabschätzung vom lokalen Detektor am Punkt i im Bild \mathcal{I} betrachtet (siehe $\mathcal{D}_i(\mathbf{x}_i; \mathcal{I})$ in Formel 4.3). Diese lokale Einschätzung an jedem Punkt der aktuellen Position des Referenzmodells werden summiert und mit dem globalen PDM verbunden. Der zu optimierende variable globale Modellparameter \mathbf{p} wird dabei durch \mathcal{R} reguliert. Als lokale Detektoren werden von Baltrusaitis u. a. [BRM16] *Conditional Neural Fields (CNFs)* [PBX09] verwendet. Nach der Optimierung des CLM wird das optimale Ergebnis

zusätzlich durch ein *Convolutional Neural Network (CNN)* validiert, um eine Abschätzung machen zu können, wie verlässlich das Ergebnis der Erkennung der Gesichtspunkte ist. Das genutzte Verfahren [BRM16] in dieser Arbeit verwendet darüber hinaus unabhängige CLMs für Augen, Lippen und Augenbraun.

Erkennung der Blickrichtung:

(Abbildung 4.3, Markierung 1.1.3)

Zur Erkennung der Blickrichtung wird das trainierte CLM der Augen verwendet. Wood u. a. [Woo+15] beschreiben, wie aus diesem Modell die Position von Augenlidern, Iris und Pupille geschätzt wird. Die geschätzte Position des Auges wird als Initialisierung für eine Berechnung der Blickrichtung verwendet, welche relativ zum Kameramittelpunkt definiert wird.

Erkennung der Kopfausrichtung/-position:

(Abbildung 4.3, Markierung 1.1.4)

Durch das verwendete CLM und dem darin enthaltenen PDM ist es möglich die Kopfausrichtung abzuschätzen. Das globale Referenzmodell ist bereits dreidimensional beschrieben. Somit kann der Drehwinkel und auch die Kopfposition direkt abgeleitet werden. Wie zuvor dargestellt, wird das Modell durch eine skalierte Orthogonalprojektion auf die Bildebene projiziert. Um die Ausrichtung zu ermitteln, wird nach der Schätzung der jeweiligen Punktposition das *n point in perspective* Problem [HR11] gelöst. Hierbei wird die Kameraposition und -ausrichtung basierend auf n dreidimensionalen Punkten eines Referenzmodells und ihren korrespondierenden zweidimensionalen Punkten in der Kameraebene geschätzt.

Normalisierung des Gesichtes:

(Abbildung 4.3, Markierung 1.1.5)

Neben den Positionen der Gesichtspunkte im Bild basiert die AU Erkennung auf einer Analyse der Textur des Gesichtes. Um die Textur für Lernalgorithmen nutzbar zu machen, muss diese standardisiert werden. Dafür beschreiben Baltrusaitis u. a. [BMR15] eine Normalisierung und Maskierung des Gesichtes. Es wird zunächst eine Transformation der geschätzten Punkte des Gesichtsmodells auf eine definierte frontale Ansicht mit neutralem Gesichtsausdruck vorgenommen. Dabei wird die *Procrustes Superimposition (PS)* verwendet. Um weitere störende Elemente zu entfernen, wird darüber hinaus die konvexe Hülle der Gesichtspunkte zur Maskierung des Gesichtes verwendet. Ein normalisiertes und maskiertes Gesicht ist in Abbildung 4.4c zu sehen.

Extraktion der Textur:

(Abbildung 4.3, Markierung 1.1.6)

Ausgehend vom normalisierten Gesicht werden erneut HOG Merkmale [DT05] (siehe Abbildung 4.4a) berechnet. Diese werden in ihrer Dimensionalität durch eine *Principal Component Analysis (Hauptkomponentenanalyse) (PCA)* reduziert, um somit eine Beschreibung der Textur zu erhalten.

Kombination der Merkmale:

(Abbildung 4.3, Markierung 1.1.7)

Sowohl die beschriebenen Texturmerkmale als auch der zuvor beschriebene variable Parameter \mathbf{p} (siehe Formel 4.2) mit den rigiden und nicht-rigiden Parametern des PDM werden zu einem Merkmalsvektor kombiniert. Dieser hat eine Größe von 1606 Dimensionen [BMR15].

Erkennung von Action Units:

(Abbildung 4.3, Markierung 1.1.8/1.1.9)

Der beschriebene kombinierte Merkmalsvektor wird verwendet, um auf annotierten Trainingsdaten das FACS zu lernen. Das verwendete Verfahren wird von Baltrusaitis u. a. [BMR15] vorgestellt. Um das Auftreten von AUs zu erkennen, werden *Support Vector Machines (SVMs)* mit linearen Kernels verwendet. In den schematischen Darstellungen dieses Kapitels werden diese Klassifikationen als „Action Units (Klassifikation)“ dargestellt. Außerdem werden *Support Vector Regressions (SVRs)* mit linearen Kernels benutzt, um die Intensität von AUs zu schätzen (siehe hierzu Baltrusaitis u. a. [BMR15]). Diese Regressionen werden mit „Action Units (Regression)“ in den Abbildungen referenziert.

Alle 17 AUs, die mit dem von mir genutzten und hier dargestellten System erkannt werden, sind in Tabelle 4.1 aufgeführt. Diese AUs dienen als Grundlage für die später diesbezüglich definierten Merkmalsgruppen. In Abbildung 4.5 sind beispielhaft unterschiedliche Gesichtsausdrücke zu sehen, die bei der Verwendung des in dieser Arbeit entwickelten Assistenzsystems auftreten.

Filterung (Abbildung 4.1, Markierung 1.2)

Bei der beschriebenen Vorverarbeitung durch das System von Baltrusaitis u. a. [BRM16] kommt es im dargestellten Szenario trotz der guten Leistung des Verfahrens dennoch zu szenariospezifischen Herausforderungen in der Erkennung.

Name	Beschreibung
AU 1	Innerer Bereich der Augenbraun ist angehoben (Inner Brow Raiser)
AU 2	Äußerer Bereich der Augenbraun ist angehoben (Outer Brow Raiser)
AU 4	Augenbraun sind gesenkt und zusammengezogen (Brow Lowerer)
AU 5	Oberes Augenlid ist angehoben (Upper Lid Raiser)
AU 6	Wange ist angehoben (Cheek Raiser)
AU 7	Augenlid ist zusammengezogen (Lid Tightener)
AU 9	Nase ist gerümpft (Nose Wrinkler)
AU 10	Obere Lippe ist angehoben (Upper Lip Raiser)
AU 12	Mundwinkel sind angehoben (Lip Corner Puller)
AU 14	Mundwinkel sind eingezogen (Dimpler)
AU 15	Mundwinkel sind herabgezogen (Lip Corner Depressor)
AU 17	Kinn ist hinaufgeschoben (Chin Raiser)
AU 20	Lippen sind nach unten und außen gezogen (Lip Stretcher)
AU 23	Lippen sind schmal nach innen zusammengepresst (Lip Tightener)
AU 25	Lippen sind geöffnet (Lip Part)
AU 26	Unterkiefer ist entspannt und Mund geöffnet (Jaw Drop)
AU 45	Blinzeln (Blink)

Tabelle 4.1.: Automatisch erkennbare Action Units:

Durch die Vorverarbeitung [BRM16] detektierbare Action Units (AUs), die die Basis der entwickelten Erkennungsverfahren sind.

Diese werden beispielhaft in Abbildung 4.6 gezeigt. Folgende Fehler in der Erkennung treten am häufigsten auf:

Erkennungsfehler 4.1 Das Gesicht wird nicht erkannt.

Erkennungsfehler 4.2 Das Gesicht wird nicht richtig erkannt.

Erkennungsfehler 4.3 Die Gesichtsmerkmale werden nicht richtig erkannt.

Erkennungsfehler 4.1 tritt besonders in Situationen auf, in denen sich die Versuchsperson aus dem Kamerabild heraus bewegt (siehe Abbildung 4.6a). Außerdem kann er auftreten, wenn die Versuchsperson ihren Kopf stark senkt, um auf das Werkstück zu schauen (siehe Abbildung 4.6b).

Erkennungsfehler 4.2 kann nur schwer an bestimmte Situationen festgemacht werden. Zumeist tritt er auf, wenn die Versuchsperson entweder texturreiche Kleidung (z. B. Karomuster) oder wie in Abbildung 4.6c einen Kragen trägt.

Erkennungsfehler 4.3 tritt besonders häufig bei starker Kopffrotation der Versuchspersonen auf (siehe Abbildung 4.6d, Abbildung 4.6e). Außerdem kann es bei Bewegung der

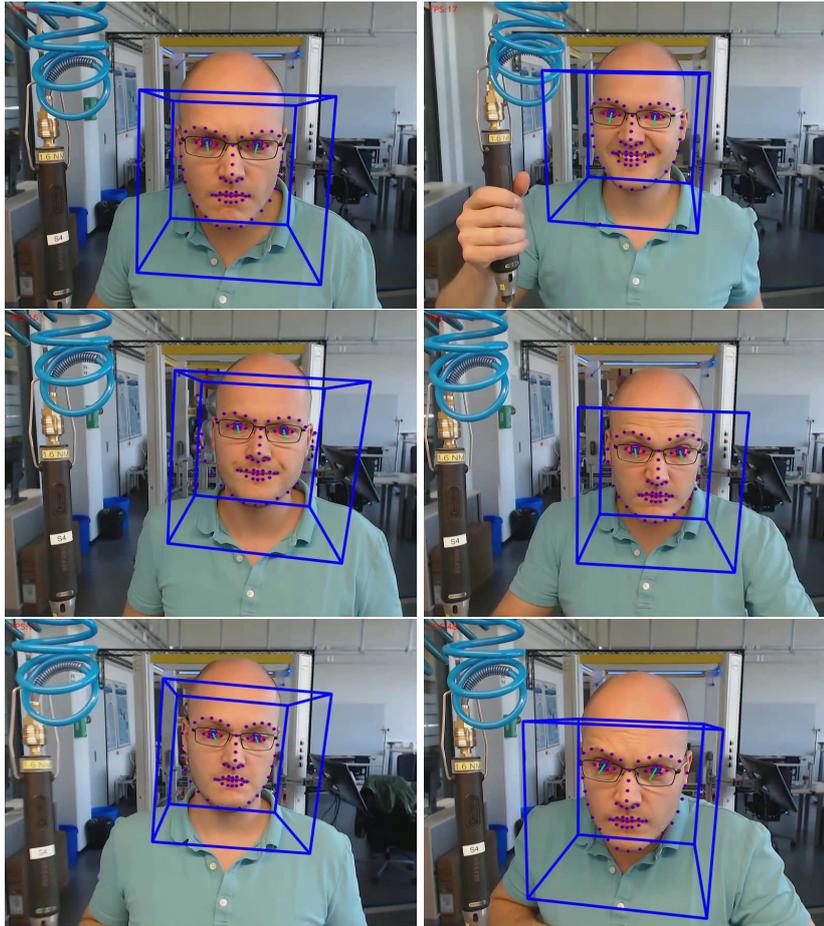
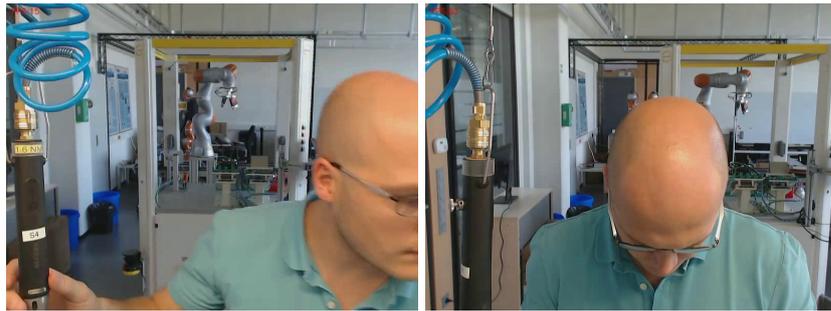


Abbildung 4.5.: Beispiele gezeigter Kopfausdrücke bei der Systeminteraktion:
 Es ist die jeweilige Visualisierung der erkannten Gesichtsmerkmale bei der Interaktion eines Probanden mit dem entwickelten Montageassistenzsystem zu sehen.

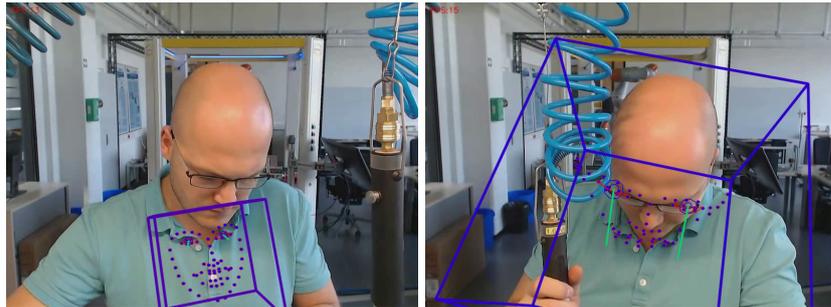
Versuchsperson aus dem Sichtfeld der Kamera (siehe Abbildung 4.6f) dazu kommen, dass das Gesicht nur teilweise sichtbar ist und die Gesichtsmerkmale dadurch falsch erkannt werden. Erkennungsfehler 4.3 kann auch bei Verdeckung des Gesichtes durch beispielsweise dem Druckluftschlauch des zu verwendenden Schraubers (siehe Abbildung 4.6g), einer Hand oder Haarstränen auftreten.

Um diese Fehler robust zu vermeiden und somit Anforderung 4.2 zu genügen, werden die Ergebnisse vor einer Weiterverarbeitung gefiltert. Für die auf den Merkmalen aufbauenden Mustererkennungsverfahren ist es besser, wenn unsichere und falsche Detektionen schon im Vorfeld gefiltert werden und somit die Gefahr einer falschen Erkennung und darausfolgenden falschen Rückschlüssen in Bezug auf die zu bewertenden Prozessschritte minimiert wird. Die Kopfausdrücke werden anhand der folgenden Merkmale gefiltert:



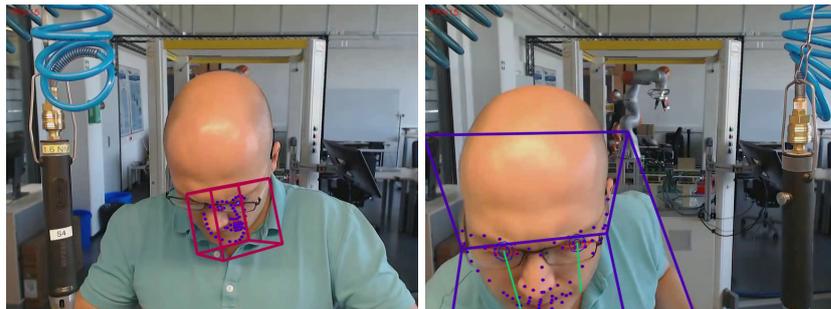
(a) Gesicht nicht erkannt

(b) Gesicht nicht erkannt



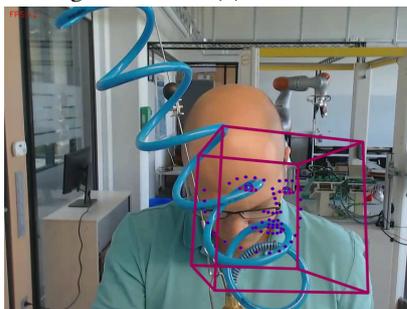
(c) Gesicht nicht richtig erkannt

(d) Merkmale nicht richtig erkannt



(e) Merkmale nicht richtig erkannt

(f) Merkmale nicht richtig erkannt



(g) Merkmale nicht richtig erkannt

Abbildung 4.6.: Auftretende Erkennungsfehler der Vorverarbeitung:

Es werden mögliche Fehler der Gesichtsmerkmalserkennung visualisiert. Das Gesicht wird nicht erkannt, das Gesicht wird nicht richtig erkannt oder die Merkmale werden nicht richtig erkannt.

Filtern anhand der Gesichtsdetektion:

(Abbildung 4.1, Markierung 1.2.1)

Erkennungsfehler 4.1 kann auf einfache Art und Weise verhindert werden. Es kann leicht ermittelt werden, ob ein Gesicht erkannt wurde oder nicht, sodass bei keiner Gesichtsdetektion auch keine Weiterverarbeitung durchgeführt wird.

Filtern anhand der Detektionssicherheit:

(Abbildung 4.1, Markierung 1.2.2)

Wie zuvor beschrieben, ist die Verarbeitungskette dazu in der Lage, eine Detektionssicherheit der Gesichtspunkte zu schätzen. Je höher diese ist, desto sicherer wurden die Gesichtsmerkmale erkannt. In Abbildung 4.6 wird die Detektionssicherheit durch die Farbe des dargestellten Quaders visualisiert. Je blauer der Quader ist, desto besser werden die Merkmale erkannt, je roter der Quader ist, desto schlechter ist die Erkennung. Durch eine Filterung anhand eines qualitativ ermittelten Schwellwertes kann die Weiterverarbeitung zusätzlich optimiert werden und Fehler wie in Abbildung 4.6e und Abbildung 4.6g werden bei der Weiterverarbeitung nicht mehr berücksichtigt. Dieser Schwellwert wurde manuell pro Versuchsperson anhand der Videoaufnahmen ermittelt.

Filtern anhand der Gesichtsgröße:

(Abbildung 4.1, Markierung 1.2.3)

Trotz der Verarbeitung der Detektionssicherheit kann es vorkommen, dass die Vorverarbeitung bei eindeutig falsch detektierten Merkmalen dennoch sicher ist, dass diese richtig erkannt wurden. Dieses ist beispielsweise in Abbildung 4.6c und Abbildung 4.6d zu sehen. Die Detektionssicherheit hat hier einen hohen Wert (siehe blauer Quader), aber die Gesichtsmerkmale werden nicht richtig erkannt. Diese Art von Fehler tritt in den meisten Fällen in Kombinationen mit einer deutlich zu kleinen (siehe Abbildung 4.6c) oder einer deutlich zu großen (siehe Abbildung 4.6d) Gesichtsdetektion auf. Aus diesem Grund wird auch die erkannte Gesichtsgröße einbezogen, um die Daten zu filtern. Anhand der vorhandenen Daten wird pro Person eine durchschnittliche Größe des Gesichtes gelernt. Falls die im aktuellen Bild ermittelte Größe zu stark von der gelernten Größe abweicht, werden die zu dem Bild gehörenden Erkennungsdaten für die weitere Verarbeitung verworfen.

4.1.2 Aufmerksamkeitserkennung

Die Aufmerksamkeitserkennung ist in Abbildung 4.1 unter Markierung 2 zu sehen. Zum besseren Verständnis wie ich den Begriff der Aufmerksamkeit in dieser Arbeit verwende, stelle ich folgende Definitionen auf:

Aufmerksamkeit Aufmerksamkeit ist die längere Konzentration des Blickes der Versuchsperson auf einen bestimmten Bereich.

Aufmerksamkeitsbereich Ein Aufmerksamkeitsbereich ist ein Bereich, dem die Versuchsperson gerade ihre Aufmerksamkeit widmet.

Die beschriebene Ermittlung der Kopfausdrücke liefert sowohl eine Schätzung der Kopfausrichtung als auch eine Schätzung der Blickrichtung. Auf dieser Basis ist eine Erkennung verschiedener Aufmerksamkeitsbereiche möglich.

Diskretisierung der Kopfausrichtung:

(Abbildung 4.1, Markierung 2.1)

Durch eine Analyse des in dieser Arbeit modellierten Prozesses wurde festgestellt, dass sich die Aufmerksamkeit in drei Bereiche unterteilen lässt. In Abbildung 4.7 sind diese Bereiche visualisiert. Sie sind aufgaben- und szenarioabhängig, können aber leicht an andere Aufgaben und Umgebungen angepasst werden.

Virtuelle Welt Dieser Bereich ist in Abbildung 4.7a und Abbildung 4.7b zu sehen. Die Versuchsperson schaut auf den jeweiligen Monitor und beschäftigt sich mit den Beschreibungen und Anweisungen des Assistenzsystems. Die Aufmerksamkeit liegt hierbei auf der virtuellen Welt.

**Reale Welt –
Werkstück** Der Bereich „Reale Welt – Werkstück“ ist in Abbildung 4.7c, Abbildung 4.7d und Abbildung 4.7e zu sehen. Die Versuchsperson widmet ihre Aufmerksamkeit in diesem Bereich dem Werkstück in der realen Welt. Die Informationen, die er durch das Assistenzsystem zuvor in der virtuellen Welt aufgenommen hat, werden nun in die reale Welt übertragen und dort am Werkstück umgesetzt.

**Reale Welt –
Sonstiges** Die Versuchsperson konzentriert sich in diesem Bereich auf andere Dinge in der realen Welt als das Werkstück. Hierbei kann es sich um Ablenkungen der Versuchsperson handeln, die beispielsweise auftreten, wenn sie sich in einer für sie unlösbaren Situation befindet und sich als letzten Ausweg an den Versuchsleiter wendet. Es

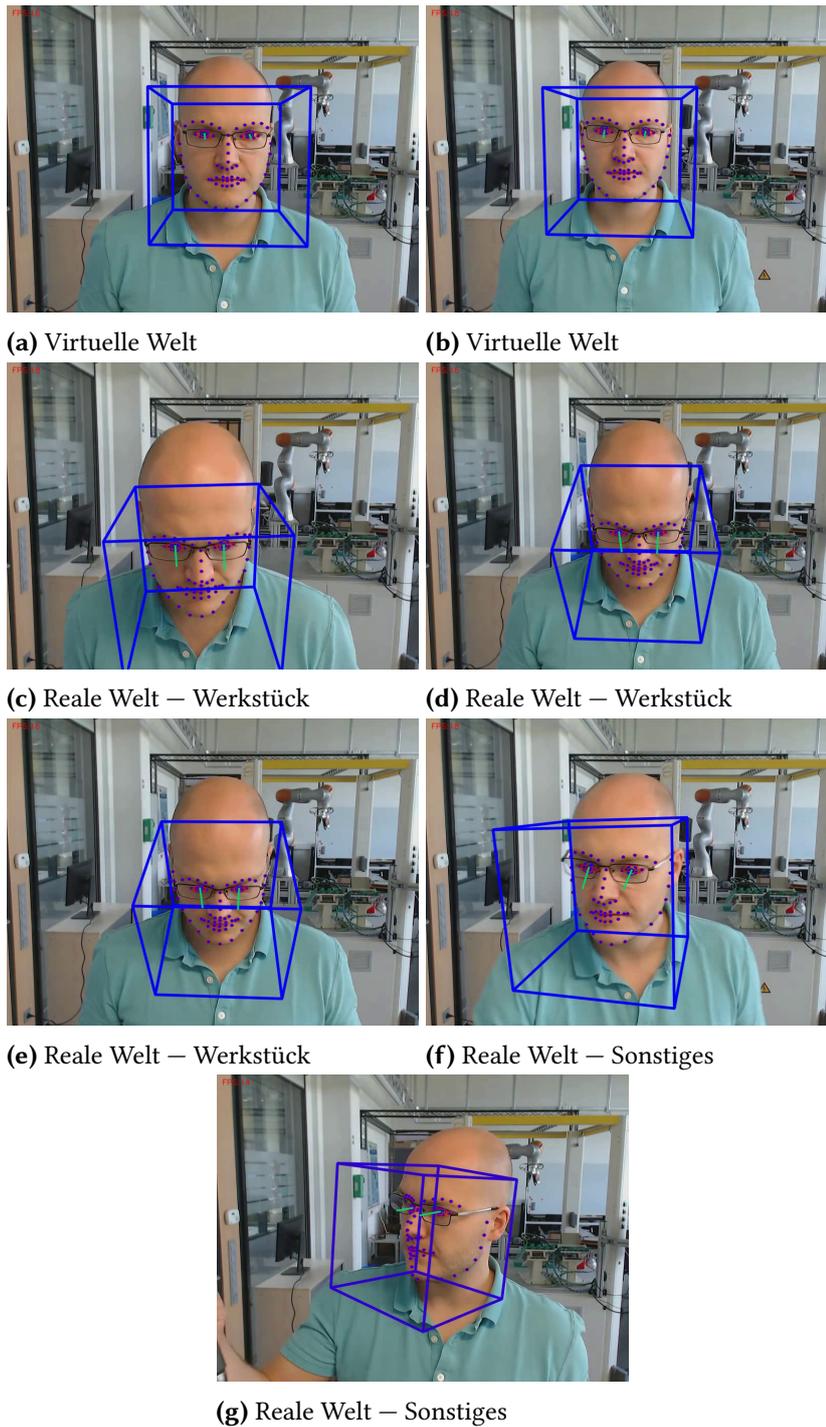


Abbildung 4.7.: Definierte Aufmerksamkeitsbereiche:

Es werden „virtuelle Welt“, „reale Welt – Werkstück“ und „reale Welt – Sonstiges“ als mögliche Aufmerksamkeitsbereiche unterschieden.

kann aber auch vorkommen, dass die Versuchsperson auf andere aufgabenrelevante Dinge wie beispielsweise einen zu verwendenden Druckluftschrauber (siehe Abbildung 4.7g) oder die Teilekästen (siehe Abbildung 4.7f) schaut.

Für eine stabile Detektion eines Aufmerksamkeitsbereiches wird die Aufmerksamkeit bewusst durch „längere Konzentration“ definiert. Diese Einschränkung muss ich machen, da die Blickrichtungserkennung in dem beschriebenen Szenario insbesondere bei schnellen Kontrollblicken zu ungenau und zu fehleranfällig ist. Um eine falsche Klassifikation im weiteren Verlauf zu vermeiden, wird bewusst darauf verzichtet, diesen Kanal zu verwenden und diese Informationen zu berücksichtigen. Bei einer stabileren Erkennung der Blickrichtung bieten gerade diese schnellen Kontrollblicke für spätere Arbeiten weitere wichtige Merkmale, um Rückschlüsse über die Prozessmodellierung zu gewinnen.

Anstatt des Blickes wird für eine robuste Erkennung der Aufmerksamkeitsbereiche die Kopfausrichtung (3 Winkel) verwendet. Diese kann robust vom beschriebenen System erkannt werden. Schnelle Blicke ohne eine Drehung des Kopfes werden dabei jedoch nicht erkannt. Eine grobe Abschätzung der Aufmerksamkeit der Versuchsperson reicht jedoch als Grundlage der später beschriebenen Verfahren aus.

Die definierten Aufmerksamkeitsbereiche sind deutlich voneinander getrennt und durch die Aufgabe und den Aufbau des Systems definiert. Somit ist für eine automatische Erkennung ein regelbasierter Ansatz die einfachste Möglichkeit der Detektion dieser Bereiche. Durch die unterschiedlichen Körpergrößen der Versuchspersonen und den somit unterschiedlichen Kopfwinkeln in Relation zur Kamera werden für jede Person individuelle Regeln aufgestellt. Basierend auf dieser Annotationsarbeit habe ich ein stabiles System entwickelt, das die drei beschriebenen Aufmerksamkeitsbereiche unterscheidet.

Eine automatische und über unterschiedliche Personen generalisierbare Erkennung möglicher Aufmerksamkeitsbereiche wäre wünschenswert. Da der Fokus dieser Arbeit jedoch nicht auf einer Schätzung der Aufmerksamkeit liegt, wurde eine solche Erkennung nicht umgesetzt.

4.1.3 Merkmalsgruppendifinition (zeitlich normiert)

Die in dieser Arbeit verwendeten und von mir definierten Merkmalsgruppen sind in Abbildung 4.1 (siehe Seite 41) unter Markierung 3 zu finden. Bei der Betrachtung von

Zeiträumen während des Lösen einer Aufgabe durch die Versuchsperson können wichtige Erkenntnisse über das Assistenzsystem gewonnen werden. Diese Arbeit hat das Ziel anhand menschlicher Verhaltensanalysen, Prozessabläufe zu analysieren und Verbesserungspotentiale aufzuzeigen. Dabei möchte ich Aufgabenschritte miteinander vergleichen, um festzustellen, ob Auffälligkeiten im menschlichen Verhalten bei der Erfüllung dieser Aufgaben erkennbar sind. Zur Herstellung dieser Vergleichbarkeit verwende ich Merkmalsgruppen, die relativ zur jeweiligen Aufgabendauer normiert werden.

Jede Merkmalsgruppe enthält dabei jeweils eine Modalität, kann aber durchaus aus mehreren einzelnen Merkmalen bestehen. Die definierten Merkmalsgruppen werde ich in den beschriebenen Ansätzen miteinander kombinieren und die unterschiedlichen Kombinationen miteinander vergleichen. Dadurch kann ich eine Aussage darüber treffen, welche Merkmalsgruppen sich am besten zur Analyse von prozessbasierten Assistenzsystemen verwenden lassen.

Zunächst eignen sich die definierten Aufmerksamkeitsbereiche sehr gut, um Erkenntnisse über die Prozessmodellierung zu erlangen. So sind besonders Fragen wie „Wie lange benötigt die Versuchsperson, um die Informationen des System zu verstehen und sich der eigentlichen Aufgabe zu widmen?“ oder „Wie oft muss die Versuchsperson die Beschreibung anschauen, um sich mit weiteren/anderen Informationen zu der Aufgabe zu versorgen?“ relevant.

Bei einer qualitativen Analyse der Studiendaten ist aufgefallen, dass ein häufiger Wechsel zwischen den einzelnen Aufmerksamkeitsbereichen auftritt, wenn die Versuchsperson die aktuelle Situation in der realen Welt nicht mit der Situation abgleichen kann, die das Assistenzsystem vorgibt. Auch die Dauer im Aufmerksamkeitsbereich „Virtuelle Welt“ steigt dann deutlich an. Um diese Variationen zu berücksichtigen, wird *Merkmalsgruppe B* (Merkmalsgruppe Blick) definiert. Diese enthält die beiden folgenden Merkmale:

Relative Aufmerksamkeitsdauer im Aufmerksamkeitsbereich „Virtuelle Welt“:

(Abbildung 4.1, Markierung 3.1)

Über einen Zeitraum wird die Relation von Aufmerksamkeitsbereich „Virtuelle Welt“ zu den Aufmerksamkeitsbereichen „Reale Welt – Werkstück“ und „Reale Welt – Sonstiges“ berechnet.

Aufmerksamkeitsbereichswechsel relativ zur Zeitraumlänge:

(Abbildung 4.1, Markierung 3.2)

Über einen Zeitraum wird die Anzahl an Wechsel zwischen den verschiedenen Aufmerksamkeitsbereichen berechnet und mit der Zeitraumlänge normalisiert.

Wie beschrieben liefert die Kopfausdruckerennung zum einen eine robuste Auftrittserkennung und zum anderen eine Abschätzung der Intensität einzelner AUs. Auch in diesem Fall werden die Merkmale anhand der betrachteten Zeitraumlänge normiert.

Auftreten jeder einzelnen AU relativ zur Zeitraumlänge:

(Abbildung 4.1, Markierung 3.3)

Über einen Zeitraum werden für jedes Einzelbild das Auftreten aller AUs unabhängig voneinander erkannt, summiert und mit der Zeitraumlänge normiert. Diese Gruppe wird im folgenden *Merkmalsgruppe AU-K* genannt und enthält 17 Merkmale (siehe Tabelle 4.1, Seite 47).

Intensität jeder einzelnen AU relativ zur Zeitraumlänge:

(Abbildung 4.1, Markierung 3.4)

Über einen Zeitraum werden für jedes Einzelbild alle AU Intensitäten unabhängig voneinander geschätzt, summiert und mit der Zeitraumlänge normiert. Auch diese Gruppe enthält 17 Merkmale (siehe Tabelle 4.1, Seite 47) und wird im folgenden *Merkmalsgruppe AU-R* genannt.

Neben der individuellen Betrachtung jeder einzelnen AU kann auch die Summe aller AUs über einen bestimmten Zeitraum bereits die nötigen Informationen enthalten, um eine Prozessschrittbewertung durchzuführen. Diese Dimensionsreduktion ist eine Schätzung der Stärke der gezeigten gesamten Mimik. Daraus werden die folgenden weiteren zu überprüfenden Merkmalsgruppen definiert:

Gesamtes AU Auftreten relativ zur Zeitraumlänge:

(Abbildung 4.1, Markierung 3.5)

Über einen Zeitraum werden für jedes Einzelbild sämtliche Auftreten der AUs klassifiziert, summiert und mit der Zeitraumlänge normalisiert. Im Folgenden wird dieses einzelne Merkmale mit *Merkmalsgruppe As-K* referenziert.

Gesamte AU Intensität relativ zur Zeitraumlänge:

(Abbildung 4.1, Markierung 3.6)

Über einen Zeitraum werden für jedes Einzelbild sämtliche AUs Intensitäten summiert

und mit der Zeitraumlänge normalisiert. Im Folgenden wird dieses einzelne Merkmal mit *Merkmalsgruppe As-R* referenziert.

4.2 Merkmalsgruppen auf Basis von System-Events

Neben der in dieser Arbeit zentralen Analyse des Kopfausdruckes können auch weitere Verhaltensweisen der Versuchspersonen in der Interaktion mit dem System einen Mehrwert für die Prozessbetrachtung bringen. Alle in diesem Abschnitt definierten Merkmale sind im folgenden in der *Merkmalsgruppe S* zusammengefasst. Am naheliegendsten ist zunächst die Betrachtung der direkten Interaktion mit dem System, d.h. eine Betrachtung, in welcher Art und Weise die nutzende Person das Assistenzsystem über die Bildschirmeingaben bedient.

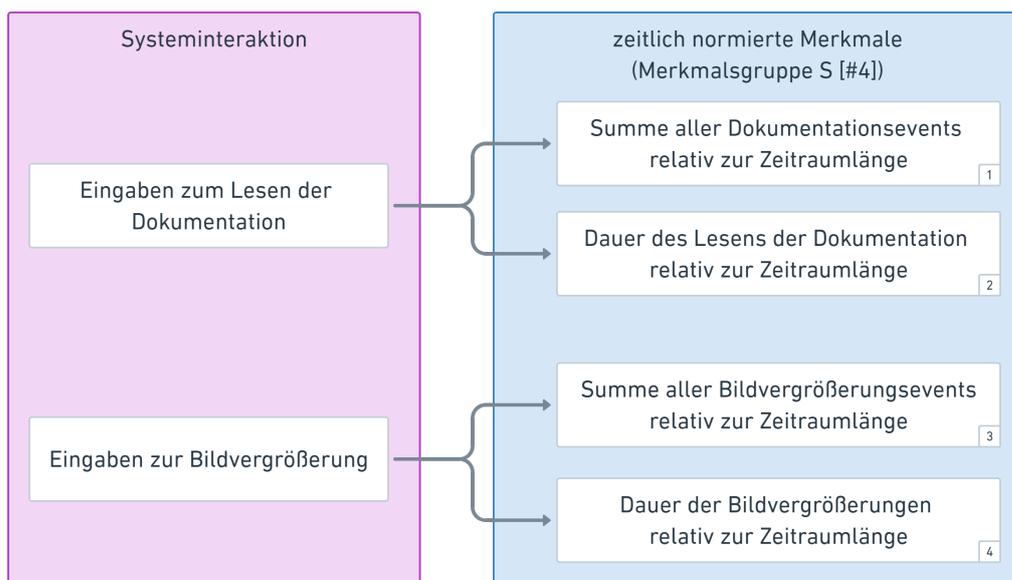


Abbildung 4.8.: Merkmalsextraktion auf Basis von System-Events:

Die System-Events bestehen aus vier Merkmalen. Markierungen in den jeweils unteren rechten Ecken der Kästen dienen zur Referenzierung im Text.

Die berücksichtigten Merkmale sind in Abbildung 4.8 dargestellt. Bei der Interaktion mit dem System ist die nutzende Person, wie bereits beschrieben, in der Lage eine Dokumentation aufzurufen und sich bereits durchgeführte Aufgabenschritte bzw. Schritte in der

Zukunft anzuschauen. Der Prozessablauf wird hierbei nicht geändert. Diese Funktion dient als reines Nachschlagewerk falls zusätzliche Informationen von der Versuchsperson benötigt werden. Sie kann dabei beliebig zu Schritten im gesamten Prozessablauf wechseln. Nach dem Nachschlagen muss sie wieder zum aktuellen Schritt zurückkehren und kann von dort aus den Prozess fortsetzen. Die folgenden beiden Merkmale beziehen sich auf die Nutzung dieses Nachschlagewerkes:

Summe aller Dokumentationsevents relativ zur Zeitraumlänge:

(Abbildung 4.8, Markierung 1)

Über einen Zeitraum werden alle Sprünge zu anderen Prozessschritten im Nachschlagewerk summiert und mit der Zeitraumlänge normalisiert.

Dauer des Lesens der Dokumentation relativ zur Zeitraumlänge:

(Abbildung 4.8, Markierung 2)

Über einen Zeitraum wird die Relation von Aufenthaltsdauer im Nachschlagewerk zur Dauer des gesamten Zeitraumes berechnet.

Neben dem Nachschlagewerk haben die Versuchspersonen zusätzlich die Möglichkeit, Bilder genauer anzuschauen. Durch ein Klick auf eine Lupe wird das Bild zunächst vergrößert. Danach besteht die Möglichkeit, mit einer Zwei-Finger-Zoomgeste das Bild weiter zu vergrößern und sich bestimmte Stellen im Bild im Detail anzuschauen.

Auch hieraus lassen sich wichtige Informationen zur Qualität der einzelnen Prozessschritte ableiten. Somit werden die folgenden weiteren Merkmale definiert:

Summe aller Bildvergrößerungsevents relativ zur Zeitraumlänge:

(Abbildung 4.8, Markierung 3)

Über einen Zeitraum werden alle Events von Bildvergrößerungen summiert und mit der Zeitraumlänge normalisiert.

Dauer der Bildvergrößerungen relativ zur Zeitraumlänge:

(Abbildung 4.8, Markierung 4)

Über einen Zeitraum wird die Relation von Anzeigedauer der vergrößerten Bilder zur Dauer des gesamten Zeitraumes berechnet.

Es wäre auch interessant, an welcher Stelle die Versuchsperson genau in einem Bild zoomt. Durch eine Analyse könnte erkannt werden, welches Detail nicht deutlich genug angezeigt wurde. Dieses Merkmal werde ich aber nicht betrachten, da die relevanten Informationen

durch das System nicht aufgezeichnet wurden. Für spätere Studien wäre eine Betrachtung der genauen Zoom-Position jedoch durchaus sinnvoll.

Insgesamt enthält Merkmalsgruppe S somit vier Merkmale der Modalität der direkten Systeminteraktion.

4.3 Merkmalsgruppe zur Berücksichtigung zeitlicher Variationen

Auch die Dauer, die die Versuchspersonen für einen Prozessschritt benötigen, kann ein relevantes Merkmal in einer Analyse der einzelnen Prozessschritte sein. Eine zeitliche Betrachtung wurde bei den bereits beschriebenen Merkmalsgruppen explizit verhindert, da diese zeitlich normiert wurden.

Dauer: Dauer eines definierten Zeitraums im Prozessablauf

Dieses Merkmal ist als einzigstes in *Merkmalsgruppe Z* vorhanden und wird berücksichtigt, um die entwickelten Methoden mit der bereits in der Praxis verwendeten reinen zeitlichen Analyse vergleichen und in einer multimodalen Analyse auch diese Informationen berücksichtigen zu können.

4.4 Zusammenfassung

Alle verwendeten Merkmale sind in Abbildung 4.9 zusammengefasst. Dabei wird auch dargestellt, welchen Ursprung die einzelnen Merkmale haben. Die beschriebenen Merkmalsgruppen werden in den Ansätzen, die in den folgenden Kapiteln vorgestellt werden, miteinander verglichen, um eine multimodale Kombination zu extrahieren, die eine Analyse des menschlichen Verhaltens am besten mit einer Prozessanalyse verbinden kann.

In diesem Kapitel habe ich zunächst die Grundlagen der entwickelten Merkmalsextraktion beschrieben. Dafür habe ich anhand der extrahierten Anforderungen ein Bildverarbeitungssystem entwickelt, das in der Lage ist, Gesichtsausdrücke und Kopfbewegungen zu nutzen, um robuste und relevante Merkmale zu erzeugen. Die durch das System erkennbaren 17 AUs bilden die für diese Arbeit relevanten Gesichtsausdrücke ab.

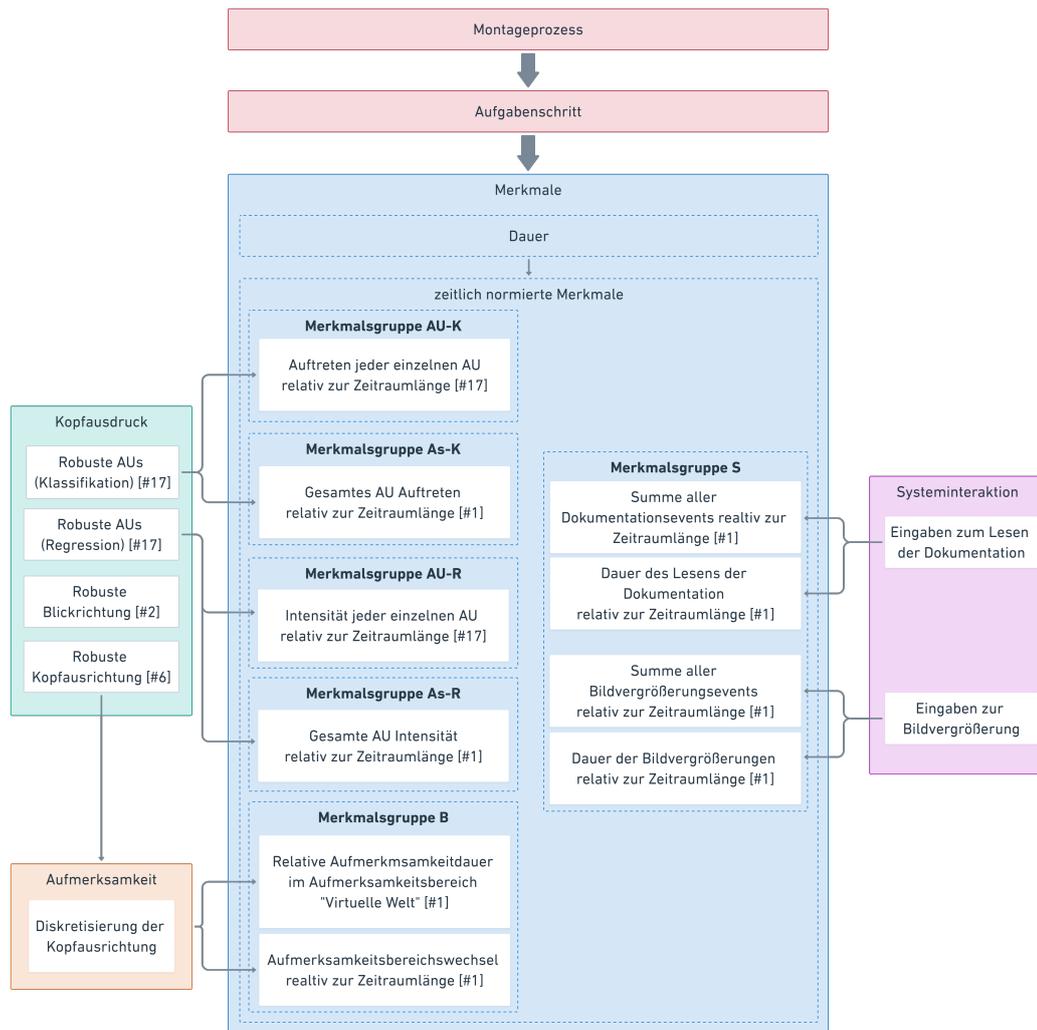


Abbildung 4.9.: Verwendete Merkmale, Merkmalsgruppen und deren Ursprung: Der Kopfausdruck, die Aufmerksamkeit, die Systeminteraktion und die Dauer der Aufgabenschritte dienen als Basis der Merkmalsgruppen. Die Anzahl der Merkmale in den einzelnen Gruppen ist in eckigen Klammern angegeben.

Die Merkmale sind durch eine zeitliche Normierung für einen Vergleich einzelner Prozessschritte definiert. Es ist mir bewusst, dass die Summierungen der einzelnen Kanäle mit anschließender Normierung relativ einfache Merkmale darstellen. Eine zeitliche Integration kann durch deutlich fortgeschrittenere Verfahren durchgeführt werden, um weitere, insbesondere zeitliche Informationen, aus den Signalen ableiten zu können. Ich zeige im Verlauf dieser Arbeit jedoch, dass auch diese einfachen Merkmale hinreichende Ergebnisse für die von mir entwickelte Analyse liefern. Ein Vergleich mit anderen komplexeren

Merkmalsgruppen auf zeitlicher Basis steht im Rahmen dieser Arbeit nicht im Mittelpunkt und kann in späteren Arbeiten durchgeführt werden.

Im weiteren Verlauf dieses Kapitels habe ich darüber hinaus Merkmalsgruppen definiert, die menschliche Verhaltensweisen bei der Interaktion mit dem Assistenzsystem abbilden. Dabei bin ich zunächst auf System-Events eingegangen um anschließend die Verwendung der Dauer als explizites Merkmal zu motivieren.

Das gezeigte System zur Merkmalserstellung bietet eine solide Grundlage auf Basis von Kopfausdrücken, um darauf aufbauend Lernverfahren zu entwickeln und zu evaluieren, die in der Lage sind, Prozessschritte miteinander zu vergleichen und somit Prozesse analysieren zu können.

Vom praktischen System zur Evaluationsplattform

„ Erfolg ist die Belohnung für schwere Arbeit.

– Sophokles

In diesem Kapitel werde ich die Entwicklung vom zuvor vorgestellten praktischen System (siehe Kapitel 3, Seite 25) zu einer Evaluationsplattform beschreiben, um zeigen zu können, ob und wenn ja wie sich eine menschliche Verhaltensanalyse zur Anwendung in der nutzerorientierten Entwicklung eignet. Darauf folgt die Beschreibung der Datenerhebung und eine Auswertung der erhobenen Fragebögen. Die Datenerhebung wird sowohl in Kapitel 6 (siehe Seite 77) als auch in Kapitel 7 (siehe Seite 105) als Korpus zum Training und zur Evaluation der entwickelten Lernverfahren verwendet.

5.1 Anpassungen des Assistenzsystems

Ich habe das von mir entwickelte produktive System (siehe Kapitel 3, Seite 25) angepasst, um folgende Eigenschaften hinzuzufügen:

- Fähigkeit zur menschlichen Verhaltensanalyse mit besonderem Fokus auf die Kopfausdrucksanalyse (siehe Kapitel 5.1.1)
- Eignung für unerfahrene nutzende Personen in einer Evaluationsumgebung (siehe Kapitel 5.1.2)
- Hervorrufen von menschlichem Verhalten durch definierte Störungen im Prozessablauf (siehe Kapitel 5.1.3)

5.1.1 Integration eines Kamerasystems

Zur Integration der Kopfausdrucksanalyse in das entwickelte Montageassistenzsystem muss auch dessen Aufbau leicht angepasst werden. Anforderung 4.1 (Ein Kamerasystem muss so platziert sein, dass der Kopf der zu beobachtenden Person so häufig und so zentral wie möglich im Kamerabild zu sehen ist.) muss dabei besonders berücksichtigt werden.

Die entwickelte Umgebung besteht aus drei nebeneinander angeordneten Arbeitsbereiche (siehe Abbildung 3.11, Seite 36), die durch einen bewegbaren Montageschlitten miteinander verbunden sind. Insgesamt hat der entwickelte Arbeitsplatz eine Breite von 2.50 Meter. Dieser Bereich kann durch eine einzelne Kamera nicht überblickt werden. Auch eine Fischaugenoptik wäre dazu nicht in der Lage, zumal diese den großen Nachteil hätte, dass besonders an Bildrändern starke Verzerrungen auftreten würden und sich der Nutzer des Weiteren nicht frontal vor der Kamera befände. Für die geforderte robuste Analyse des Kopfausdrucks (Anforderung 4.2) ist dieses unbedingt zu vermeiden.



Abbildung 5.1.: Kamerapositionen im Montageassistenzsystem:

In jedem Arbeitsbereich ist eine Kamera (schematisch dargestellt und rot eingekreist) oberhalb der Monitore befestigt, die die Versuchsperson möglichst zentral aufnimmt.

Aus diesem Grund habe ich ein Multikamera-System zur Videoaufnahme implementiert, indem jeweils eine Kamera am oberen Rand die Bildschirme der einzelnen Arbeitsbereiche angebracht wurde (siehe Abbildung 5.1). Dadurch kann Anforderung 4.1 erfüllt und auch die Voraussetzungen zur Erfüllung von Anforderung 4.2 geschaffen werden.

Die Position oberhalb des Monitors wurde gewählt, da der Fokus auf der Evaluation des verwendeten Assistenzsystems liegt, welches sich in den Bildschirmen manifestiert. Eine Position unterhalb der Monitore, sodass man einen besseren Blickwinkel auf das Gesicht der arbeitenden Person hat, wenn diese etwas am Werkstück befestigt, wurde verworfen, da hierbei sowohl beim Blick auf die Monitore, als auch beim Blick auf das Werkstück starke

Rotationen in Relation zur Kameraposition auftreten. Bei der gewählten Position oberhalb der Monitore tritt nur eine starke Rotation auf, wenn die Person auf das Werkstück schaut. Beim Blick auf den Bildschirm wird Anforderung 4.1 dennoch erfüllt. Gerade der Zeitraum des Lesens und Verstehens des Arbeitsschrittes erachte ich als besonders wichtig in der Evaluation des Prozesses, da schon zu diesem Zeitpunkt bei auftretenden Problemen eine Reaktion im Kopfausdruck zu erwarten ist.

5.1.2 Anpassung an unerfahrene Versuchspersonen

Durch eine Vorstudie mit drei unerfahrenen Versuchspersonen ist deutlich geworden, dass die komplette Durchführung des modellierten Montageprozesses zu lange dauert, um eine angestrebte Durchlaufdauer von etwa einer Stunde zu erreichen. Somit habe ich mich entschieden, die letzte Arbeitsstation nicht in der Datenerhebung zu berücksichtigen und bereits nach der zweiten Station die Durchführung der Studie zu beenden. Die durchzuführenden Aufgaben wurden dadurch auf 87 reduziert (von vormals 116 (siehe Tabelle 3.1, Seite 32)) und der zeitliche Horizont konnte eingehalten werden.

Außerdem wurde ein druckluftbetriebenes Presswerkzeug im Montageassistenzsystem durch eine Rohrzange ersetzt. Die Versuchspersonen sollen nicht unnötig durch die Benutzung unbekannter Werkzeugs belastet und abgelenkt werden. Darüber hinaus soll die anfängliche Erklärung des Assistenzsystems durch den Versuchsleiter so kurz und kompakt wie möglich sein. Mithilfe der Rohrzange kann die betreffende Aufgabe adäquat und vergleichbar erfüllt werden.

5.1.3 Manuelle Einführung von Störungen im Prozessablauf

Bei dem durch die nutzerorientierte Entwicklung erstelltem Montagesystem ist das Risiko zu hoch, dass bei der durchgeführten Datenerhebung keine Störungen und Schwierigkeiten im Prozessablauf auftreten. Das Ziel dieser Arbeit ist jedoch, genau diese Art der Störungen mit menschlichen Verhaltensweisen in Verbindung zu bringen, um darauf aufbauend die Prozesse bewerten und Potentiale im Prozess aufzeigen zu können. Besonders für überwachte Klassifikationsverfahren sind *Ground Truth* Daten zum Training essentiell. Auch für eine Evaluation der gelernten Verfahren werden diese Daten benötigt. Darum

wurden künstliche Prozessstörungen eingeführt und der entwickelte Prozess so angepasst, dass eine definierte Anzahl dieser Störungen abgebildet wird. Diese Störungen lösen ein natürliches menschliches Verhalten während der Interaktion mit dem System aus. Somit kann eine standardisierte und evaluierbare Umgebung geschaffen werden, um den Forschungsfragen nachzugehen.

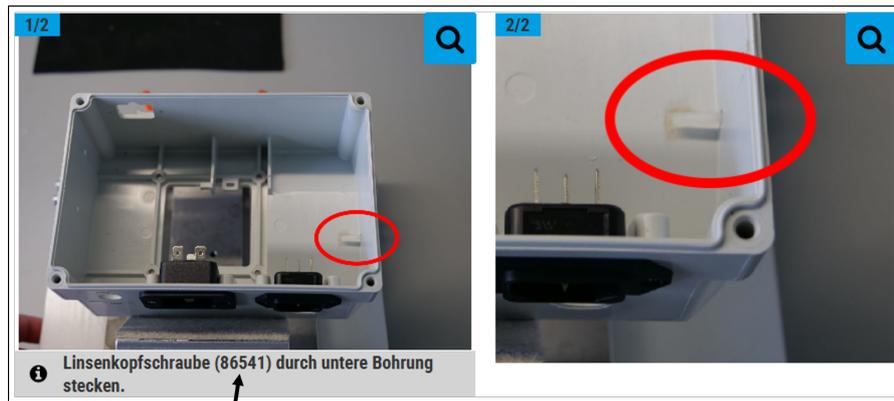
Einerseits werden schwerwiegende, andererseits leichte Störungen im Prozessablauf integriert. Schwerwiegende Störungen lösen in jedem Fall eine Reaktion im Verhalten der Versuchspersonen aus, da der Prozess ohne eine Auflösung der Störung nicht fortgesetzt werden kann. Leichte Störungen dienen vielmehr der Verunsicherung der Versuchsperson und der Steigerung der Komplexität des Aufgabenschrittes. Diese Störungen müssen nicht explizit gelöst werden und werden zumeist nicht bewusst durch die Versuchsperson erkannt. Im Folgenden werde ich diese Störungen definieren und die gemachten Anpassungen am Prozess darstellen.

Schwerwiegende Störung Als schwerwiegende Prozessstörungen werden Störungen in einzelnen Prozessschritten bezeichnet, die nur sehr schwer durch Interaktion mit dem Assistenzsystem aufgelöst werden können.

Die erste schwerwiegende Prozessstörung tritt im 8. Montageschritt auf. Die Versuchspersonen haben sich bis zu diesem Schritt mit der Bedienung des Systems vertraut gemacht und sind in der Lage, es mit einer angemessenen Geschwindigkeit zu steuern. Wie in Abbildung 5.2 zu sehen ist, müssen die Versuchspersonen eine Linsenkopfschraube durch eine Bohrung stecken. Die benötigten Teile sind in Kästen, die mit der jeweiligen Artikelnummer beschriftet sind, zu finden. In diesem Fall ist jedoch die angegebene Artikelnummer falsch und existiert nicht in der realen Welt. Diese Störung kann durch die Versuchsperson auf zwei Wege gelöst werden. Zum einen kann sie die Information ignorieren und schließt vom dargestellten Bild direkt auf das vorhandene Teil und wählt das ähnlichste aus den zur Verfügung stehenden Teilen aus. Zum anderen kann die Versuchsperson im Nachschlagewerk nach anderen Prozessschritten suchen, die auch dieses Teil verwenden, um dort die richtige Artikelnummer zu finden.

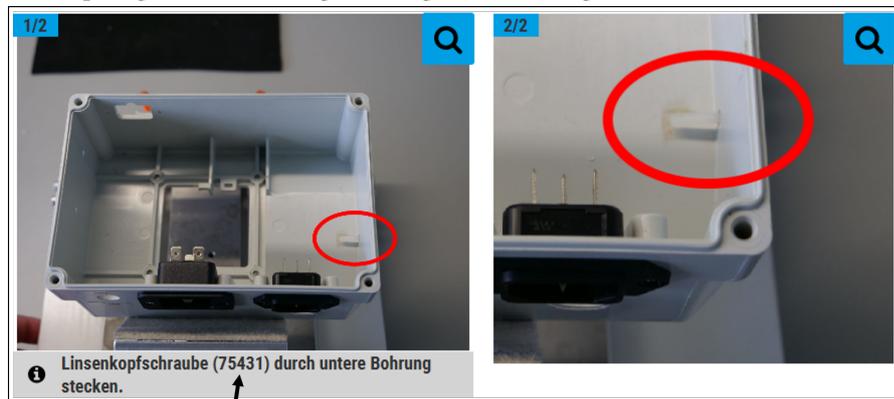
Die gleiche Art der Prozessstörung tritt nochmals in Schritt 36 auf. Auch hier ist die Artikelnummer falsch angegeben. Die Lösungsstrategien, wie bei der zuvor beschriebenen Störung in Schritt 8, treffen auch hier zu.

Ein weiterer schwerwiegender Fehler wurde im 35. Schritt implementiert. Die richtige und die fehlerhafte Anweisungen sind in Abbildung 5.3 zu sehen. Es wurde die wichtige



Richtige Artikelnummer

(a) Ursprüngliche Anweisung der Aufgabe mit richtiger Artikelnummer



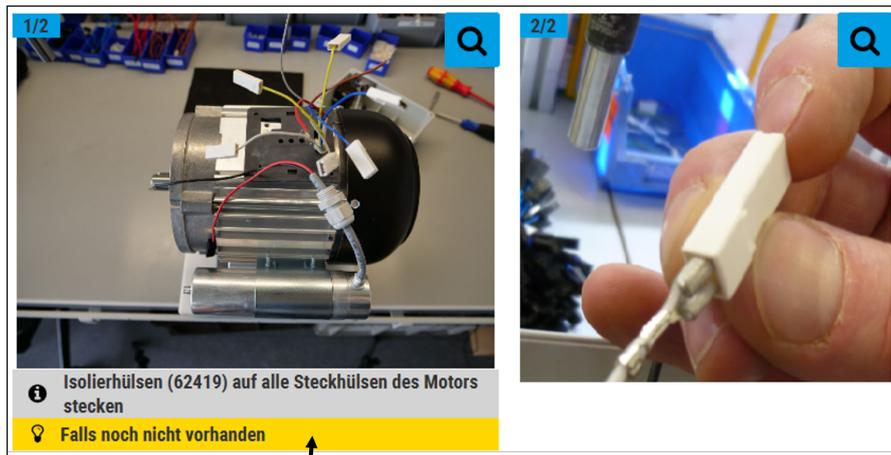
Falsche Artikelnummer

(b) Angepasste Anweisung der Aufgabe mit falscher Artikelnummer

Abbildung 5.2.: Prozessstörung im 8. Schritt des Montageprozesses:

In dieser manuell hinzugefügten, schwerwiegenden Prozessstörung wird eine nicht vorhandene Artikelnummer eines Montageteils referenziert.

Information „Falls noch nicht vorhanden“ entfernt. Diese Änderung erscheint nicht gravierend, die Versuchspersonen stehen dadurch aber vor einer großen Herausforderung. Sie werden dazu aufgefordert, sechs Isolierhülsen zu verwenden und zu befestigen. Dieses ist aber nicht möglich, da sie bereits vorhanden und montiert sind. Das jeweils linke Bild zeigt den Motor bereits mit den angebrachten Hülsen, das Detailbild kann aber durch die Versuchsperson nicht nachvollzogen werden. Somit macht der Prozessschritt keinen Sinn mehr. Die Versuchsperson hat hier nur eine Möglichkeit der Auflösung dieser Störung. Sie muss den Schritt ignorieren und einfach ohne etwas in der realen Welt zu ändern zum nächsten Schritt übergehen.



(a) Ursprüngliche Anweisung der Aufgabe mit Hinweis

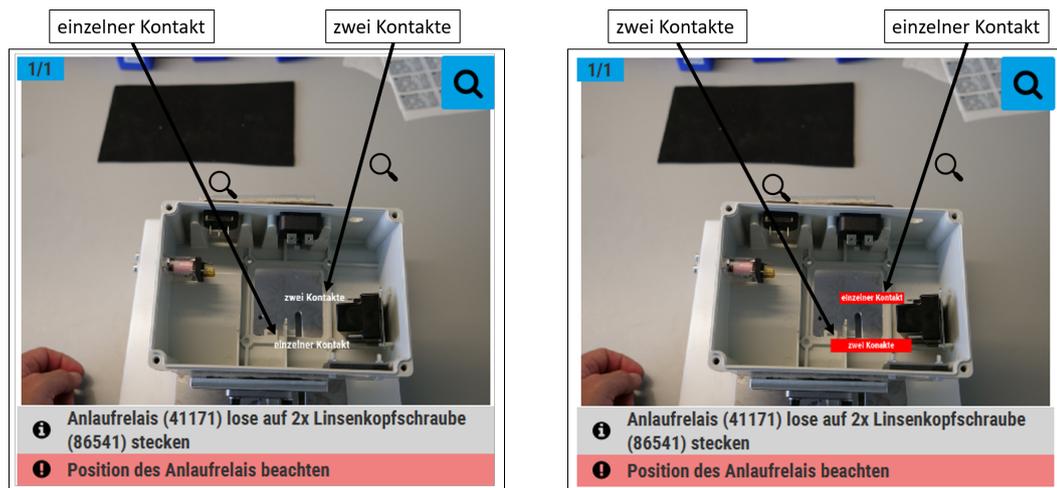


(b) Angepasste Anweisung der Aufgabe ohne Hinweis

Abbildung 5.3.: Prozessstörung im 35. Schritt des Montageprozesses:

In dieser manuell hinzugefügten, schwerwiegenden Prozessstörung wird ein wichtiger Hinweis zur aktuellen Aufgabe nicht gegeben.

Abbildung 5.4 zeigt die nächste modellierte Prozessstörung. Diese basiert auf einem absichtlich eingeführten Montagefehler im 23. Schritt. In diesem Schritt sollen die Versuchspersonen ein Anlaufrelais einbauen. Das Relais hat auf der einen Seite zwei Kontakte und auf der anderen Seite einen Kontakt. Die richtige Montage ist in Abbildung 5.4a zu sehen. Der Versuchsperson wurde jedoch die Anweisung aus Abbildung 5.4b präsentiert. Somit bauen sie unwissentlich das Anlaufrelais mit einer falschen Orientierung ein. Diese falsche Orientierung stört den Prozess in den meisten Fällen im 58. Prozessschritt (siehe Abbildung 5.5). Dabei soll das falsch eingebaute Relais mit Kabeln verbunden werden. Den



(a) Ursprüngliche Anweisung der Aufgabe mit richtiger Annotation

(b) Angepasste Anweisung der Aufgabe mit vertauschter Annotation

Abbildung 5.4.: Prozessstörung im 23. Schritt des Montageprozesses:

In dieser manuell hinzugefügten, schwerwiegenden Prozessstörung wird ein Montageteil durch die Versuchspersonen unwissentlich mit einer falschen Orientierung verbaut. Die Annotationen („einzelner Kontakt“, „zwei Kontakte“) sind zur Verdeutlichung in der Abbildung hervorgehoben.

Versuchspersonen fällt dabei auf, dass die aktuelle Situation und der aktuelle Aufbau in der realen Welt nicht mit dem gezeigten Bild übereinstimmen. Der rot markierte Anschluss (siehe Abbildung 5.5) ist zwar vorhanden, es ist aber ein weiterer Anschluss auf der Seite verwendbar. In seltenen Fällen, fällt der Fehler erst in Schritt 60 (siehe Abbildung 5.8b, Seite 71) auf, da dort der untere Kontakt des Anlaufrelais explizit referenziert wird.

Die Versuchsperson kann die Störung nur auflösen, wenn sie erkennt, dass das betroffene Teil falsch eingebaut wurde. Dafür muss sie im Nachschlagewerk den Einbauschnitt zunächst finden und mit dem aktuellen Schritt abgleichen. Nach der Realisierung des Fehlers hat sie zwei Möglichkeiten die Situation aufzulösen. Die Versuchsperson kann das betroffene Teil falsch herum eingebaut lassen und es auf der anderen Seite anzuschließen. Diese Art der Reparatur ist jedoch unwahrscheinlich, da der Versuchsperson vorher klar gemacht wurde, dass sie die Anleitung so genau wie möglich zu verfolgen hat, um den nötigen Qualitätsstandard zu erreichen. Ein falsch eingebautes Teil widerspricht der Nähe zur Anleitung. Eine wahrscheinlichere und erwartete Lösung ist, dass die Versuchsperson den gemachten Fehler realisiert, die Anleitung aus Schritt 23 (siehe Abbildung 5.4b) ignoriert und entgegengesetzt zu dieser Beschreibung das betreffende Relais ausbaut und



Abbildung 5.5.: Schritt 58, in dem der Fehler aus Schritt 23 realisiert wird:

In Schritt 58 wird die Prozessstörung aus Schritt 23 realisiert. Das verbaute Teil soll nun angeschlossen werden und die Beschreibung passt nicht zur aktuellen Montagesituation.

richtig erneut einbaut. Somit repariert sie den zuvor gemachten Fehler und kann mit dem dargestellten Schritt 58 (siehe Abbildung 5.5) fortfahren.



Abbildung 5.6.: Prozessstörung im 84. Schritt des Montageprozesses:

In dieser manuell hinzugefügten, schwerwiegenden Prozessstörung wird eine nicht lösbare Sichtprüfung gefordert.

Die letzte modellierte schwerwiegende Störung ist im letzten von der Versuchsperson durchzuführenden Schritt implementiert. Dieser Schritt ist in Abbildung 5.6 zu sehen und verlangt eine finale Sichtprüfung. Die Versuchsperson muss durch eine Sichtkontrolle sicherstellen, dass die zuvor durchgeführten 52 Schritte korrekt durchgeführt wurden. Der Prozessschritt ist logisch richtig, er wird jedoch durch ein einzelnes Bild dargestellt (siehe Abbildung 5.6). Mit dieser Information kann keine sinnvolle Sichtprüfung von 52

durchgeführten Schritten erfolgen. Diese Störung wäre nur durch ein erneutes Durchgehen und gleichzeitiger Kontrolle der betreffenden Schritte im Nachschlagewerk adäquat lösbar. Diese Strategie ist aber bei der durchgeführten Erhebung unwahrscheinlich, da sich die Versuchspersonen zu diesem Zeitpunkt bereits etwa 60 Minuten in der Evaluationssituation befinden und von einer sinkenden Motivation auszugehen ist die Qualität zu sichern. Deutlich wahrscheinlicher ist, dass die Versuchspersonen Irritation zeigen und versuchen, möglichst gut mit dem dargestellten Bild die Montage zu überprüfen und dann die Sichtprüfung als erfolgreich abschließen.

Jede dieser beschriebenen Prozessstörungen kann auch durch den Einbezug des Versuchsleiters gelöst werden. Die Versuchspersonen wurden jedoch instruiert, dass dieses nur der letzte mögliche Schritt ist und wenn möglich vermieden werden sollte.

Neben den schwerwiegenden Störungen wurden auch leichte Störungen in den Prozess integriert.

Leichte Störung Als leichte Prozessstörungen werden Störungen in einzelnen Prozessschritten bezeichnet, die einfach durch direkte Interaktion mit dem Assistenzsystem aufgelöst werden können oder keiner Auflösung bedürfen.

Diese Störungen sollen den Prozessablauf nicht gravierend unterbrechen, sondern dienen der Verunsicherung der Versuchspersonen und der möglichen Erhöhung der kognitiven Anforderung durch schwierigere und unklarere Aufgabenbeschreibungen. Durch diese leichten Störungen, die die Versuchsperson zumeist nicht bewusst wahrnimmt, sollen besonders unterbewusste Verhaltensreaktionen hervorgerufen werden.

Die erste leichte Störung ist in Schritt 11 modelliert (siehe Abbildung 5.7). Nachdem im vorherigen Schritt eine Unterlegscheibe auf die Linsenkopfschraube aufgebracht wurde, soll das Relais dort zwischen dieser Scheibe und der Kastenwand aufgesetzt werden. Die leichte Störung besteht darin, dass der Hinweis, wo genau das Relais eingesetzt werden soll, entfernt wurde. Somit muss die Versuchsperson selber die Entscheidung der Relaispositionierung treffen und es entsteht das Risiko, dass sie das Relais falsch einbaut.

Eine weitere leichte Störung wurde in Schritt 60 eingeführt (siehe Abbildung 5.8). Hierbei wurden wie zuvor Informationen entfernt, die der Klarheit dienen. In diesem Fall wurde ein Detailbild und der Hinweis entfernt, dass „egal welche“ Litze aus dem Motor verwendet werden kann. Auch dieses kann zu Verwirrung führen, da die Versuchsperson entscheiden muss, welche Litze sie benutzt.



Wichtiger Hinweis

(a) Ursprüngliche Anweisung der Aufgabe mit Hinweis



Kein Hinweis

(b) Angepasste Anweisung der Aufgabe ohne Hinweis

Abbildung 5.7.: Prozessstörung im 11. Schritt des Montageprozesses:

In dieser manuell hinzugefügten, leichten Prozessstörung wird ein wichtiger Hinweis zur aktuellen Aufgabe nicht gegeben.

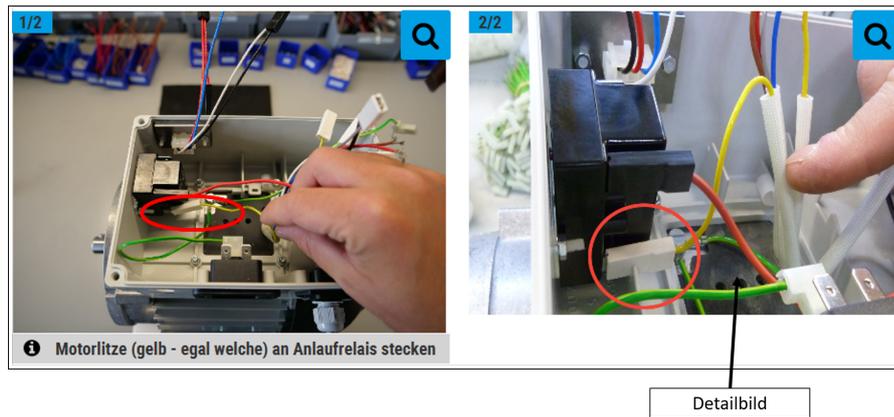
Auch in Schritt 62 (siehe Abbildung 5.9) wurden Informationen entfernt. Dabei fehlt wiederum das Detailbild und die Anweisung, an welche Position vom Relais die Litze angeschlossen werden muss. Dieses kann durch eine genau Betrachtung und manuellem Vergrößern des Bildes ermittelt werden, ist aber dennoch schwer zu erkennen.

In der letzten modellierten Störung im Schritt 63 wurde erneut ein Hinweis entfernt. In diesem Fall wurde der Tipp entfernt, dass man ein störenden Kondensatoranschluss biegen soll, um einen Stecker besser anbringen zu können. Ohne diesen Tipp ist das Anbringen des Steckers deutlich schwieriger.

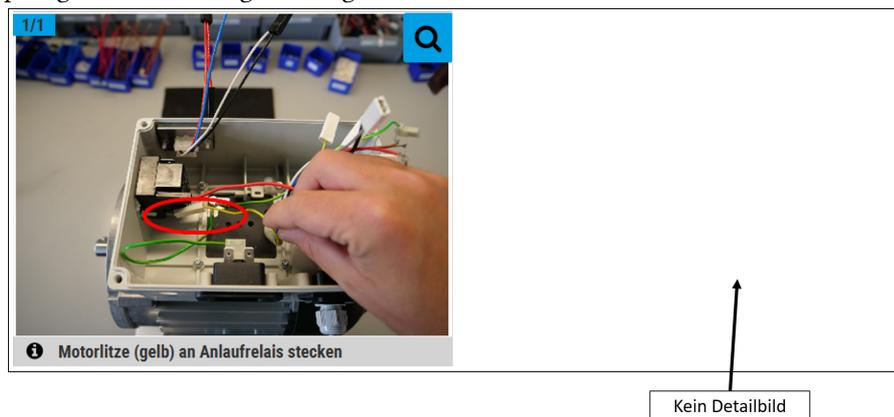
Insgesamt wurden somit 5 schwerwiegende und 4 leichte Prozessstörungen modelliert und in den vorhandenen Prozess integriert.

5.2 Beschreibung der Datenerhebung

Die Erhebung der Daten wurde durch die Ethik-Kommission der Universität Bielefeld genehmigt (siehe Anhang A.1.1, Seite 139). Die erhobenen Daten werden sowohl im entwickelten überwachten Verfahren (siehe Kapitel 6, Seite 77) als auch im unüberwachten



(a) Ursprüngliche Anweisung der Aufgabe mit Detailbild



(b) Angepasste Anweisung der Aufgabe ohne Detailbild

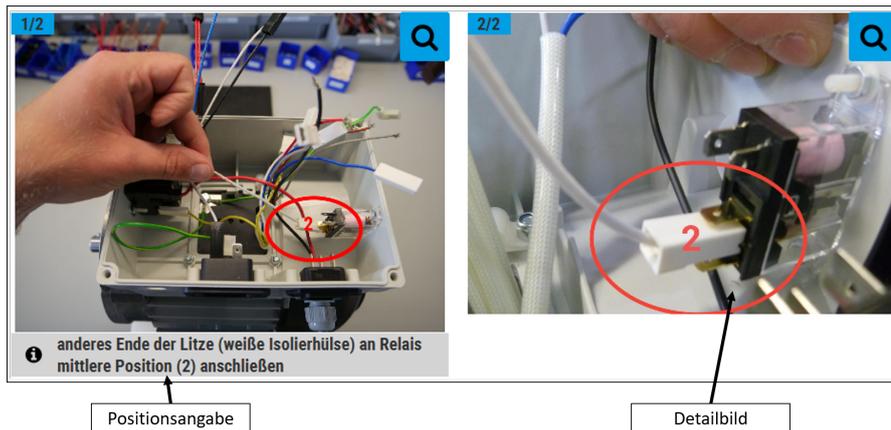
Abbildung 5.8.: Prozessstörung im 60. Schritt des Montageprozesses:

In dieser manuell hinzugefügten, leichten Prozessstörung wird ein Detailbild zur aktuellen Aufgabe nicht gezeigt.

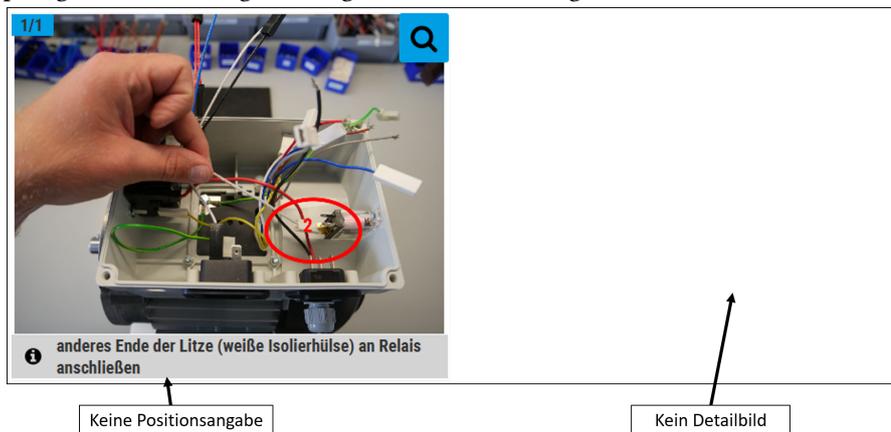
Verfahren (siehe Kapitel 7, Seite 105) verwendet. Da ich mich auf die Analyse unerfahrener Personen konzentriere und somit keinerlei Vorkenntnisse benötigt wurden, konnten durch Aushänge in der Universität Bielefeld und in der Fachhochschule Bielefeld Versuchspersonen geworben werden. Sie haben für die Durchführung der Erhebung eine Aufwandsentschädigung von 10€ erhalten.

Wie in Tabelle 5.1 dargestellt, umfasst die betrachtete Stichprobe 15 Personen, wovon 60 Prozent männlich sind. Das Durchschnittsalter beträgt 32,6 Jahre. Die jüngste Person ist 21, die älteste Person 60 Jahre alt.

Vor Beginn des Erhebung wurden die einzelnen Versuchspersonen mit dem Assistenzsystem vertraut gemacht. Um die Durchführung nicht durch unterschiedliche Erklärungen des Projektleiters zu beeinflussen, wurde ein standardisierter Ablauf festgelegt, um die



(a) Ursprüngliche Anweisung der Aufgabe mit Positionsangabe und Detailbild



(b) Angepasste Anweisung der Aufgabe ohne Positionsangabe und ohne Detailbild

Abbildung 5.9.: Prozessstörung im 62. Schritt des Montageprozesses:

In dieser manuell hinzugefügten, leichten Prozessstörung wird ein Detailbild und eine Positionsangabe zur aktuellen Aufgabe nicht gezeigt.

Bedienung mit Hilfe des Touch-Displays, die Funktionen der Oberfläche (wie das Aufrufen der Hilfeseiten und das Zoomen in die Beschreibungsbilder) und die Verwendung der im System integrierten Druckluftschrauber zu erklären. Nachdem die Teilnehmenden die Einverständniserklärung (siehe Anhang A.1.2, Seite 140) gelesen und unterschrieben hatten, wurde mit der Datenerhebung begonnen. Der Versuchsleiter befand sich aus Sicherheitsgründen weiterhin im Raum. In der Einführung wurde den Versuchspersonen jedoch deutlich gemacht, dass dieser nur im äußersten Notfall Hilfestellung gibt und sie die Aufgaben alleine in der Interaktion mit dem System lösen müssen.

Nach der Durchführung der Datenerhebung haben die Versuchspersonen einen SUS Fragebogen (siehe Anhang A.1.3, Seite 141), einen RTLX Fragebogen (siehe Anhang A.1.4,

Kategorie	Wert
Umfang [N]	15
Geschlecht [% männlich]	60%
Alter [M (SD)]	32,6 (11,61)

Tabelle 5.1.: Teilnehmendendaten der Datenerhebung:

Es sind allgemeine Informationen zu den Versuchspersonen dargestellt.

Seite 142) und einen Persönlichkeitsfragebogen [Ram+12] (siehe Anhang A.1.5, Seite 143) ausgefüllt.

Aus dem Korpus mussten zwei Personen entfernt werden, da es in beiden Fällen zu einem Absturz von Systemteilen gekommen ist. Im ersten Fall wurde die Kameraaufnahme unerwartet beendet, im zweiten Fall ist ein Fehler in der Prozesssteuerung aufgetreten. Daraus folgt ein zu verwendender Korpus, der 13 Personen enthält. Außerdem wurde von einer Person der RTLX Fragebogen nicht ausgefüllt.

Des Weiteren ist bei einer manuellen Analyse des Datensatzes aufgefallen, dass die als schwerwiegend eingeführte Störung in Schritt 36 keine Reaktion im Verhalten der Versuchspersonen auslöst. Es handelt sich um das zweite Auftreten einer falschen Artikelnummer. Ich vermute, dass die Versuchspersonen bereits bei der ersten falschen Artikelnummer die Möglichkeit eines Auftretens eines solchen Fehlers gelernt haben und nicht mehr die Artikelnummer verwenden, um die relevanten Teile zu finden. Dieser Aufgabenschritt wird durch die mangelnde Reaktion der Versuchsperson nicht als Prozessstörung angesehen.

5.3 Auswertung der Fragebögen

Der aus den erhobenen Fragebögen resultierende SUS ist in Tabelle 5.2 zu sehen. Bei einer Betrachtung aller Nutzer kommt es zu einem SUS von etwa 83. Dieses ist nach der Einteilung von Brooke [Bro13] ein *gutes* Ergebnis (Schulnote B). Die Nutzer wurden auch nach ihrer Erfahrung in der „Nutzung von Montageassistenzsystemen“ befragt (Skala 1–6). Betrachtet man nun die 6 Versuchspersonen mit einer hohen Montageassistenzerfahrung (≥ 4), wird ein SUS von 92 erreicht. Dieses ist nach Brooke [Bro13] *exzellent* (Schulnote A). Der SUS von 77 bei Personen mit einer geringen Montageassistenzsystemerfahrung (< 4) ist nach Brooke [Bro13] auch als *gutes* Ergebnis einzuordnen. Es ist erstaunlich, dass

trotz der integrierten schwerwiegenden und leichten Prozessstörungen die beschriebenen guten Werte der Usability abgeleitet werden können.

Gruppe	Anzahl	SUS (0–100)
Gesamte Stichprobe	15	83
Erfahrung Montageassistenzsysteme ≥ 4	6	92
Erfahrung Montageassistenzsysteme < 4	9	77

Tabelle 5.2.: System Usability Score (SUS) des entwickelten Assistenzsystems:

Es ist der SUS der gesamten Stichprobe, aller Versuchspersonen mit hoher Montageassistenzsystemerfahrung und aller Versuchspersonen mit geringer Montageassistenzsystemerfahrung dargestellt (gerundet auf ganze Zahlen).

Der Unterschied zwischen den Gruppen kann durch eine unterschiedliche Erwartungshaltung gegenüber dem Montagesystem begründet werden. Personen, die bereits viel Erfahrung im Umgang mit Montageassistenzsystemen in der Praxis erlangt haben, können das entwickelte System mit den ihnen bekannten Systemen vergleichen. Der hohe SUS lässt darauf schließen, dass das Assistenzsystem mindestens genauso gut, wenn nicht sogar besser in diesem Vergleich abschneidet. Versuchspersonen mit wenig Erfahrung im Umgang mit Montagesystemen können diesen Vergleich nicht machen und bewerten aus diesem Grund das System schlechter, da sie es gegebenenfalls mit Alltagskategorien wie Smartphones vergleichen, die anscheinend eine bessere Usability haben.

Dimension	Wert (0–100)
Geistige Anforderung	61.04
Körperliche Anforderung	43.5
Zeitliche Anforderung	32.71
Leistung	35.83
Anstrengung	47.92
Frustration	45.83
RTLX Wert	44.47

Tabelle 5.3.: Raw Task Load Index (RTLX) Auswertung der Datenerhebung:

Es werden die einzelne Dimensionen und das gesamte Ergebnis der RTLX Auswertung der Datenerhebung dargestellt.

Die einzelnen Dimensionen der RTLX Erhebung und der ermittelte RTLX Wert sind in Tabelle 5.3 dargestellt. Der RTLX lässt Rückschlüsse auf die kognitive Belastung zu (siehe Abschnitt 2.2.1, Seite 12). Ein Wert von 44.47 auf einer Skala von 0 bis 100 ist relativ

niedrig und lässt auf eine durchschnittliche kognitive Belastung schließen. Es fällt jedoch auf, dass die Dimension *geistige Anforderung* mit einem Wert von 61.04 im Vergleich zu den anderen Dimensionen deutlich erhöht ist. Diese Erhöhung ist durch die Art der Aufgabe erklärbar. Komplexe, zuvor noch nicht gesehene Schritte zum Zusammenbau eines unbekanntes Werkstückes müssen von der Beschreibung auf die aktuelle Situation übertragen werden. Diese Aufgabe ist kognitiv fordernd und wird durch den erhöhten Wert bestätigt. Außerdem haben die Versuchspersonen die Aufgabe zum ersten Mal durchgeführt. Auch dieser Neuheitseffekt kann zu der hohen geistigen Anforderung führen. Der Effekt wird durch die neuartige Arbeitsumgebung zusätzlich erhöht. Schließlich können auch die integrierten Prozessstörungen diesen höheren Wert unterstützen.

Nach der Erhebung der Fragebögen hatten die Versuchspersonen die Möglichkeit, weitere Verbesserungen, die sich aus störenden Interaktionen mit dem System ergeben haben, in einem Interview zu benennen. Es wurden zunächst allgemeine Störungen wie der Arbeitsplatzaufbau („Schläuche des Schraubers befinden sich immer vor den Bildschirmen.“, „Teileboxen sind zu weit weg.“) oder Bedienschwierigkeiten sowohl mit dem System („Zoom ist zu träge.“) als auch bei der Montage („Motor ist zu schwer“) genannt. In Hinblick auf den Prozessablauf wurden von den Versuchspersonen durchgehend die integrierten schwerwiegenden Störungen referenziert. Dabei wurde immer zuerst der falsche Einbau des Relais beschrieben und die daraus resultierende nötige Reparatur. Darauf folgend kam es häufig zu der Beschreibung des ersten Auftretens der falschen Artikelnummer und des Schrittes mit den bereits vorhandenen Steckhülsen. Seltener wurde die nicht durchführbare Sichtprüfung beschrieben. Auf sonstige Prozessschritte ist keine der Versuchspersonen eingegangen. Die modellierten Störungen sind so stark, dass sie andere möglicherweise erkannte Probleme überdecken.

5.4 Zusammenfassung

Dieses Kapitel beschreibt die Überführung des praktischen Systems in eine Evaluationsumgebung. Dabei wird zunächst das entwickelte Kamerasystem vorgestellt und die Anpassung der Studienlänge an unerfahrene Versuchspersonen beschrieben. Hervorzuheben ist die darauffolgende manuelle Einführung von Prozessstörungen zur Evaluation und zum Training der in den folgenden Kapiteln beschriebenen Verfahren. Es wurden einerseits schwerwiegende, andererseits leichte Prozessstörungen integriert. Diese praxisrelevanten Störungen lösen häufig vorkommende bewusste und unbewusste Reaktionen im Verhalten

der Versuchspersonen bei der Benutzung eines prozessbasierten Montageassistenzsystems aus. Dieses Verhalten dient als Grundlage für die entwickelten Erkennungsverfahren. Die qualitative Analyse der Leistung der gelernten Verfahren im Bezug auf die restlichen und nicht manuell angepassten Schritte ist dabei besonders interessant, da weitere Potentiale zur Prozessverbesserung aufgedeckt werden können. Die Verfahren lernen zwar manuell hinzugefügte Fehler, die darauffolgende Analyse konzentriert sich aber auf die nicht angepassten und bereits im Vorfeld durch die nutzerorientierte Entwicklung erstellten Prozessschritte.

Dieses Kapitel schließt mit der Beschreibung der Datenerhebung und der Auswertung der erhobenen Fragebögen. Ich konnte zeigen, dass das entwickelte System eine gute und bei erfahrenen Nutzern sogar eine exzellente Nutzbarkeit aufweist. Der hohe Wert der Nutzbarkeit, gerade bei erfahrenen Personen, lässt eine deutliche schlechtere Nutzbarkeit bei aktuell verbreiteten und in der Praxis genutzten Montageassistenzsystemen vermuten. Außerdem konnte ich durch die Erhebung des RTLX eine Schätzung der kognitiven Belastung vornehmen. Es fällt auf, dass besonders die Dimension der geistigen Belastung einen hohen Wert erzielt.

Erkennung von Prozessstörungen durch überwachte Verfahren

“ *If a machine is expected to be infallible, it cannot also be intelligent.*

– Alan M. Turing

Zur Erkennung von Störungen im Prozess und somit verbesserungswürdigen Prozessschritten entwickle ich in diesem Kapitel einen überwachten maschinellen Lernansatz. Ich modelliere diese Erkennung als ein binäres Entscheidungsproblem. Als positive Klasse werden die beschriebenen Aufgabenschritte festgelegt, in denen es zu schwerwiegenden Prozessstörungen (siehe Abschnitt 5.1.3, Seite 63) kommt bzw. die schwerwiegende Prozessstörung erkannt wird. In der negativen Klasse sind alle anderen Aufgabenschritte enthalten. Somit evaluiere ich mithilfe der in Abschnitt 5.2 (siehe Seite 70) beschriebenen Datenbasis, ob es möglich ist, diese schwerwiegenden Fehler anhand menschlicher Verhaltensweisen gegenüber allen anderen Schritten zu erkennen. Zusammenfassend definiere ich die folgenden Hypothesen:

- Hypothese 6.1** Aufgabenschritte mit schwerwiegenden Prozessstörungen lassen sich erkennen.
- Hypothese 6.2** Basierend auf einer Auftrittserkennung von AUs lassen sich bessere Ergebnisse erzielen als auf einer Intensitätserkennung der AUs.
- Hypothese 6.3** Eine Betrachtung einer kondensierten Repräsentation des Gesichtsausdruckes in einem einzelnen Merkmal reicht zur Erkennung von schwerwiegenden Störungen nicht aus.

Hypothese 6.4 Ein kombiniertes System mehrerer multimodaler Merkmalsgruppen liefert die besten Ergebnisse bei der Erkennung von schwerwiegenden Störungen.

6.1 Systembeschreibung

In diesem Abschnitt beschreibe ich detailliert das von mir entwickelte Klassifikationssystem. Mit dessen Hilfe werde ich die genannten Hypothesen durch unterschiedliche Kombinationen von Klassifikatoren, Vorverarbeitungsschritten und Merkmalsgruppenkombinationen untersuchen. Ein Überblick des entwickelten Systems ist in Abbildung 6.1 zu sehen.

Die verwendeten Daten bestehen aus allen Aufgabenschritten der 13 verwendbaren Probanden der beschriebenen Erhebung. Somit sind in der positiven Klasse 52 und in der negativen Klasse 1079 Aufgabenschritte enthalten und es herrscht ein großes Ungleichgewicht zwischen den Klassen. Die verwendeten Klassifikatoren sind jedoch in der Lage, ein Ungleichgewicht in den Klassen zu berücksichtigen.

Für ein Training und eine Evaluation wird der Korpus zunächst zufällig gruppenbasiert anhand der Personen in *Trainingsdaten* und *Testdaten* geteilt (siehe Abbildung 6.1, oben). Die Trainingsdaten enthalten sieben und die Testdaten sechs Personen. Dieses Teilungsverfahren wurde gewählt, um zu verhindern, dass Personen sowohl im Datensatz zur Optimierung als auch im Datensatz zur Validierung vorhanden sind. Dadurch kann eine objektivere Evaluation im Hinblick auf die Personenunabhängigkeit des entwickelten Verfahrens durchgeführt werden.

6.1.1 Optimierung

Die Optimierung ist in Abbildung 6.1 unter Markierung 1 zu sehen. Um unterschiedliche Klassifikatoren mit den jeweils besten Parametern und den am besten geeigneten Merkmalsgruppen zu vergleichen, wird das *Tree-based Pipeline Optimization Tool (TPOT)*¹ [Ols+16] verwendet. Mit TPOT kann die Erstellung von Verarbeitungsketten von maschinellen Lernansätzen automatisiert werden. Somit ist weder ein manuelles *Grid Search*

¹<http://epistasislab.github.io/tpot/>

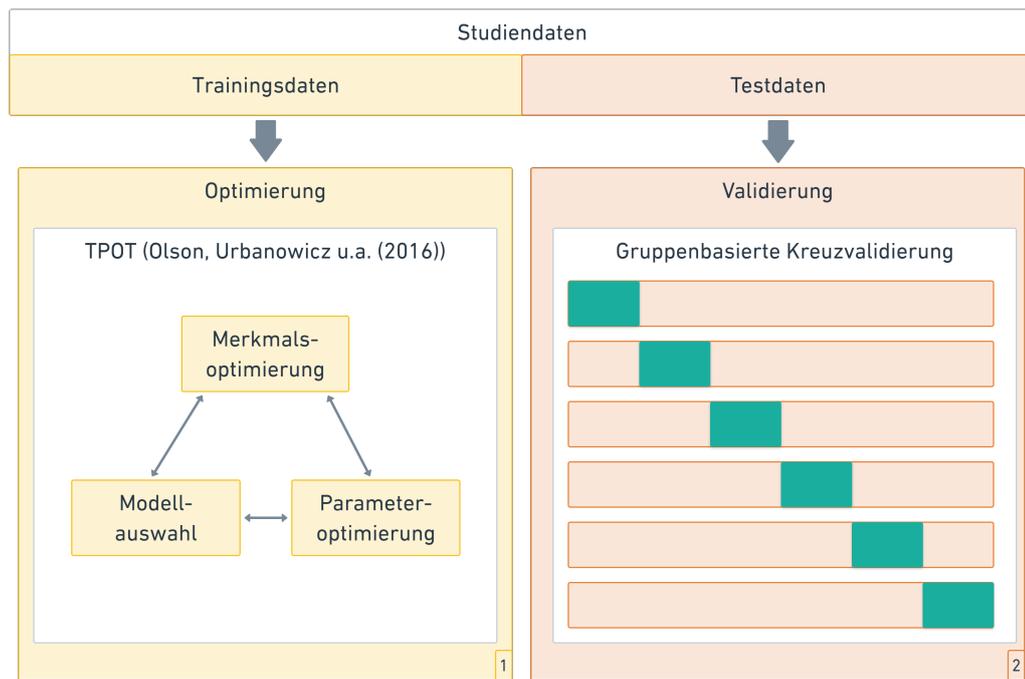


Abbildung 6.1.: Überwachtes Klassifikationssystem:

Es ist das Trainings- und Testverfahren des Klassifikationssystems schematisch dargestellt. Die Studiendaten werden in Trainings- und Testgruppen unterteilt. Die Trainingsdaten werden für eine Optimierung der Klassifikationskette verwendet [Ols+16] und die Testdaten für die Validierung der optimierten Klassifikationsketten mithilfe einer gruppenbasierten Kreuzvalidierung. Die rot-grünen Balken in der Validierung stellen die einzelnen Faltungen der Kreuzvalidierung dar (grün=Test, rot=Training). Markierungen in den jeweils unteren rechten Ecken der Kästen dienen zur Referenzierung im Text.

nach optimalen Parametern noch eine manuelle Auswahl der besten Vorverarbeitungsschritte nötig. TPOT basiert auf den im *scikit-learn Framework*² [Ped+11] implementierten Algorithmen und kann daraus durch baumbasierte Optimierungsverfahren eine optimale Klassifikatorverarbeitungskette ableiten. Mithilfe von evolutionären Algorithmen (in diesem Fall genetische Algorithmen) und festgelegten Optimierungskriterien ist TPOT in der Lage durch *Mutationen* die Vielzahl an möglichen Kombinationen der Verarbeitungskette zu evaluieren. Dabei werden Merkmale, Parameter und Modelle optimiert.

Um verschiedene Klassifikationsverfahren miteinander vergleichen zu können, werden Verfahren auf Basis von *Entscheidungsbäumen*, *Boosting*, *Nachbarschaften* und *Stützvektoren* unabhängig voneinander betrachtet und mithilfe von TPOT optimiert. Dabei kann

²<http://scikit-learn.org>

in der Optimierung jeweils aus einer Fülle von Vorverarbeitungsschritten zur Merkmals-optimierung gewählt werden. Diese werde ich im Folgenden kurz darstellen. Für weitere Informationen verweise ich auf die jeweils genannten Publikationen bzw. der Beschreibung im scikit-learn Framework.

Merkmalsoptimierung

Jede Optimierung, die im Training durchgeführt wird, kann auf die in diesem Abschnitt beschriebenen Vorverarbeitungsmöglichkeiten zurückgreifen. Ich habe eine Sammlung aus allgemein üblichen Vorverarbeitungsschritten gewählt. Eine Auswahl musste getroffen werden, da andernfalls die Rechenzeit der Optimierung mit den vorhandenen Ressourcen nicht mehr praktisch händelbar wäre. TPOT wählt dabei selbständig die Kombination für ein optimales Klassifikationsergebnis aus. Zu den Beschreibungen gebe ich jeweils auch die verwendeten Funktionen in scikit-learn oder TPOT an.

Binarisierung (sklearn.preprocessing.Binarizer)

Anhand eines Schwellwerts wird jedes Merkmal binarisiert. Der Schwellwert wird dabei in der Parameteroptimierung angepasst.

Minimum-Maximum Skalierung (sklearn.preprocessing.MinMaxScaler)

Jedes Merkmal wird im Bereich 0 bis 1 skaliert.

Maximum-Absolut Skalierung (sklearn.preprocessing.MaxAbsScaler)

Jedes Merkmal wird mit dem jeweiligen absoluten Maximalwert skaliert.

Naive Skalierung (sklearn.preprocessing.StandardScaler)

Jedes Merkmal wird durch Entfernen des Mittelwertes und Skalierung auf Varianz 1 standardisiert.

Robuste Skalierung (sklearn.preprocessing.RobustScaler)

Um eine Robustheit zu erreichen, wird jedes Merkmal durch Entfernung des Medians und Skalierung mithilfe des *Interquartilsabstandes* standardisiert. Dieser Abstand beinhaltet die mittleren 50 Prozent der Daten. Die Randbereiche und somit die Ausreißer werden bei der Skalierung nicht beachtet.

Normalisierung (sklearn.preprocessing.Normalizer)

Alle Merkmale eines Datenpunktes werden normalisiert. Durch die Parameteroptimierung wird die verwendete Norm angepasst.

Summe aller Nullen (`tpot.builtins.ZeroCount`)

Die Anzahl an Nullen und an Nicht-Nullen im jeweiligen Merkmalsvektor wird als neues Merkmal hinzugefügt.

Merkmalsverschmelzung (`sklearn.cluster.FeatureAgglomeration`)

Mithilfe einer Clusteranalyse werden Merkmale, die ähnlich zueinander sind, zusammengefasst. Dabei wird die Distanz zwischen den Clustern und die gewählte Distanzmetrik in der Parameteroptimierung angepasst.

Hauptkomponentenanalyse (`sklearn.decomposition.PCA`)

Es wird eine PCA auf den Merkmalen mit einer zufallsbasierten Singulärwertzerlegung durchgeführt. Dabei wird in der Parameteroptimierung die Konfiguration der Singulärwertzerlegung angepasst.

Radiale Basisfunktionen (`sklearn.kernel_approximation.RBFSampler`)

Es wird eine Kernel-Approximation der Merkmale mithilfe des *Random Kitchen Sinks* Verfahren erstellt, um eine approximierete Abbildung der radialen Basisfunktionskernel zu liefern. Die Konfiguration des Kernels wird in der Parameteroptimierung angepasst.

Varianz zur Merkmalsauswahl (`sklearn.feature_selection.VarianceThreshold`)

Die Merkmale werden anhand ihrer Varianz gefiltert. Ist diese unter einem bestimmten Schwellwert, wird das Merkmal nicht berücksichtigt. Der Schwellwert wird in der Parameteroptimierung angepasst.

Statistik zur Merkmalsauswahl (`sklearn.feature_selection.SelectFwe`, `sklearn.feature_selection.SelectPercentile`)

Basierend auf statistischen Tests werden die besten Merkmale ausgewählt. Die Art der Berechnung und die Konfiguration der Verfahren werden durch die Parameteroptimierung bestimmt.

Modellauswahl

Im Folgenden stelle ich die Klassifikationsmodelle kurz vor, die für die Optimierung verwendet wurden. Dabei wurde jeweils eine eigene Verarbeitungskette für jedes einzelne Modell erstellt, um diese in der folgenden Auswertung miteinander vergleichen zu können.

Entscheidungsbäume (sklearn.ensemble.RandomForestClassifier, sklearn.ensemble.ExtraTreesClassifier)

Bei der Klassifikation auf Basis von Entscheidungsbäumen werden die verbreiteten Verfahren *Random Forrest* [Bre01] und *Extra Tree* [GEW06] verwendet. Diese basieren auf sogenannten *Wäldern* von Entscheidungsbäumen, welche durch randomisierte Arten des Trainings erstellt werden. Ich werde im Folgenden diese Klassifikationsmodelle auch als *Bäume* referenzieren. Dabei wird zum einen die jeweiligen Konfigurationen der Wälder in der Parameteroptimierung angepasst, zum anderen wird das verwendete Modell gewählt.

Boosting (sklearn.ensemble.GradientBoostingClassifier)

Durch eine Kombination von *schwachen Lernern* wird durch das *Gradient Boosting* Verfahren [Fri01] ein Klassifikator erstellt. Hierbei wird in der Parameteroptimierung die Lernrate und die Konfiguration der schwachen Lerner angepasst.

Nachbarschaften (sklearn.neighbors.KNeighborsClassifier)

Nachbarschaftsverfahren basieren auf einer Betrachtung der direkten Umgebung zu einem Datenpunkt [CD07]. Dabei wird in der Parameteroptimierung die Anzahl der zu verwendenden Nachbarn, die Gewichtung der einzelnen Nachbarn und die zu verwendende Distanzmetrik variiert. Im weiteren Verlauf dieser Arbeit wird dieses Modell auch als *Nachbar* referenziert.

Stützvektoren (sklearn.svm.SVC)

Bei Stützvektorverfahren wird die Klassengrenze zwischen Klassen gelernt und durch Stützvektoren abgebildet. Dabei wird in der Parameteroptimierung zwischen linearen Methoden und radialem Basisfunktionskernel gewählt. Außerdem werden die Parameter der Verfahren optimiert. Diese Art der Klassifizierung wird beispielsweise von Steinwart und Christmann [SC08] und Hearst u. a. [Hea+98] näher erläutert und im Folgenden als *SVM* referenziert.

Parameteroptimierung

Zur weiteren Leistungssteigerung der resultierenden Klassifikationskette wird durch die TPOT Umgebung zusätzlich eine Optimierung der Konfigurationsparameter der einzelnen Kettenelemente durchgeführt. Diese habe ich bereits bei den jeweiligen Elementen im Abschnitt zuvor beschrieben. Dabei wird zunächst für die einzelnen der zuvor beschriebenen

Bausteine definiert, in welchen Bereichen die jeweiligen spezifischen Konfigurationsparameter automatisch angepasst werden. Die Optimierung führt TPOT automatisch anhand der definierten Konfiguration und im Wechselspiel mit der Modellauswahl und der Merkmalsoptimierung durch (siehe Abbildung 6.1).

6.1.2 Validierung

Die Validierung ist in Abbildung 6.1 (siehe Seite 79) unter Markierung 2 zu sehen. Zur Evaluation der zuvor optimierten Klassifikationsketten werden die beschriebenen Testdaten verwendet. Dabei wird die *Leave One Group Out Kreuzvalidierung (LGOCV)* benutzt, da jeder Proband sowohl die Aufgaben der positiven zu lernenden Klasse als auch die Aufgaben der negativen Klasse durchgeführt hat. Somit wäre bei einer vollständigen randomisierten Auswahl der Training- und Testmenge keine Unabhängigkeit gegeben und die Ergebnisse würden verfälscht. Bei der LGOCV werden Validierungsdaten anhand einer Gruppenzugehörigkeit getrennt. In der hier beschriebenen Evaluation nehme ich eine Trennung anhand der Versuchspersonen vor. Somit stellt jede Versuchsperson eine einzelne Gruppe dar. Da sechs Personen in den Validierungsdaten enthalten sind, können sechs Faltungen durchgeführt werden. Pro Faltung wird eine Person als Testmenge verwendet und das Training der zuvor durch die Optimierung ermittelte besten Verarbeitungskette erfolgt auf den übrigen fünf Probanden. Die Ergebnisse der Faltungen werden gemittelt und bestimmen die Güte des Klassifikatorsystems.

Ermittlung der Klassifikationsgüte

Um die verschiedenen Verarbeitungsketten der Klassifikation vergleichen zu können, verwende ich die Größe der *Area under Receiver-Operating-Characteristic Curve (AUROC)* [Bra97; Faw06; Pow11]. Um dieses verbreitete Gütekriterium nutzen zu können, erstelle ich jeweils ein *Receiver-Operating-Characteristic (ROC)* Diagramm. Hierbei wird die *Richtig-Positiv-Rate (RPR)* gegen die *Falsch-Positiv-Rate (FPR)* aufgetragen, wobei die Entscheidungsschwelle des Klassifikators variiert wird. Dieses hat den großen Vorteil, dass man keine Aussagen der Güte relativ zu einem bestimmten Schwellwert treffen muss, sondern alle Schwellwerte des Klassifikationsmodell einbezogen werden. Dadurch kann eine allgemeingültige Aussage zum gewählten Modell getroffen werden.

Richtig-Positiv-Rate (RPR)

Als RPR, Sensitivität oder Trefferquote wird die Relation zwischen positiven Klassifikationen bei betrachteten positiven Beispielen r_p (richtig positiv) zu allen tatsächlich positiven Beispielen $r_p + f_n$ (richtig positiv + falsch negativ) bezeichnet:

$$P(\text{pos. Klassifikation} \mid \text{tatsächlich pos.}) = \frac{r_p}{r_p + f_n}$$

Anders gesagt ist die RPR die Wahrscheinlichkeit, dass bei einer positiven Klassifikation das Beispiel auch tatsächlich zu der positiven Klasse gehört.

Falsch-Positiv-Rate (FPR)

Als FPR oder auch Ausfallrate wird die Relation zwischen positiven Klassifikationen bei tatsächlich negativen Beispielen f_p (falsch positiv) zu allen tatsächlich negativen Beispielen $r_n + f_p$ bezeichnet:

$$P(\text{pos. Klassifikation} \mid \text{tatsächlich neg.}) = \frac{f_p}{r_n + f_p}$$

Anders gesagt ist die FPR die Wahrscheinlichkeit, dass bei einer positiven Klassifikation das Beispiel tatsächlich zu der negativen Klasse gehört, also ein Fehlalarm ist.

Wie auch Fawcett [Faw06] beschreibt, werden im ROC Diagramm (Beispiele sind in Anhang A.2 (siehe Seite 146) zu sehen) somit der Nutzen (RPR) gegen die Kosten (FPR) eines Klassifikators aufgetragen. Jeder einzelne Punkt in dem dargestellten Raum repräsentiert eine Konfiguration der betrachteten Verarbeitungskette der Klassifikation. Bei einer FPR von 0 und einer RPR von 0 gibt die betrachtete Klassifikationskette nie eine positive Antwort zurück und somit wird nie ein falsch positiver Fehler begangen. Im Gegensatz dazu wird aber auch niemals eine richtig positive Klassifikation durchgeführt. Bei einer FPR von 1 und einer RPR von 1 kommt es zu einem genau gegenteiligen Verhalten. Hierbei gibt die betrachtete Klassifikationskette immer die positive Klasse zurück. Somit kommt es auch in allen möglichen Fällen zu falsch positiven Fehlern. Beide Klassifikationsmodelle sind denkbar schlecht. Ein optimaler Klassifikator ist bei einer FPR von 0 und einer RPR von 1 gegeben. Dabei kommt es zu keinem Fehlalarm und einer perfekten Klassifikation der tatsächlich richtigen Beispiele. Je näher eine Klassifikator Konfiguration diesen Punkt kommt, desto besser ist sie.

Als Richtwert wird bei ROC Diagrammen oftmals eine zufällige Entscheidung angenommen. Die von mir in dieser Arbeit dargestellten ROC Diagramme (siehe Anhang A.2,

Seite 146) enthalten jeweils diese zufällige Entscheidung als Referenz (siehe dargestellte Diagonale $x = y$). Wenn ein Klassifikator in 25 Prozent aller Fälle die positive Klasse zufällig wählt, kann angenommen werden, dass 25 Prozent der positiven Klasse richtig erkannt werden und auch 25 Prozent der negativen Klasse als positiv deklariert werden. Wählt ein Klassifikator in 90 Prozent aller Fälle die positive Klasse, kann angenommen werden, dass 90 Prozent der positiven Klasse richtig erkannt werden und auch 90 Prozent der negativen Klasse als positiv erkannt werden. Somit entsteht die Diagonale im ROC Diagramm. Ein Ergebnis unterhalb der Diagonalen (schlechtere Leistung als eine zufällige Entscheidung) kann durch eine Negation der Ausgabe des Klassifikators zu einem positiven Ergebnis gewandelt werden.

Die oben beschriebenen und in der Optimierung verwendeten Klassifikationsmodelle sind in der Lage eine Wahrscheinlichkeit der trainierten Klassen zu schätzen. Um nun ein vollständiges ROC Diagramm zu erstellen, wird die Schwelle für eine Zuordnung zu einer Klasse anhand dieser Wahrscheinlichkeiten festgelegt und für das Diagramm variiert.

Für ein Vergleichsmaß unterschiedlicher Klassifikationsmodelle wird nun diese zweidimensionale Repräsentation der Güte in einem Diagramm zu einem einzelnen Wert gewandelt. Hierfür wird die Größe des Bereichs unterhalb der ROC berechnet (auch AUROC genannt). Je größer der Bereich ist, je besser ist die Leistung der gewählten Verarbeitungskette zur Klassifikation. Da eine zufällige Entscheidung die beschriebene Diagonale erzeugt (Größe des Bereichs unter der Diagonalen ist 0.5) und Klassifikatoren unterhalb dieser Diagonalen wie beschrieben gewandelt werden können, sollte ein Klassifikator stets einen AUROC-Wert größer als 0.5 erreichen.

Um die verwendete LGOCV in ein ROC Diagramm zu überführen, wird eine Mittelung der einzelnen Faltungen durchgeführt [Faw06]. Somit entsteht pro Klassifikationsmodell ein einzelnes Diagramm und der AUROC-Wert kann verglichen werden.

Zur Abschätzung der Güte eines Klassifikators hat sich eine grobe Einteilung des AUROC-Wertes etabliert:

- 0.9–1.0 = ausgezeichnet (engl. excellent), Schulnote 1 (engl. A)
- 0.8–0.9 = gut (engl. good), Schulnote 2 (engl. B)
- 0.7–0.8 = befriedigend (engl. fair), Schulnote 3 (engl. C)
- 0.6–0.7 = schlecht (engl. poor), Schulnote 4 (engl. D)
- 0.5–0.6 = fehlgeschlagen (engl. fail), Schulnote 5 (engl. E)

Verfahren zur statistischen Vergleichbarkeit

Um die definierten Hypothesen bewerten zu können, wird eine detaillierte statistische Auswertung durchgeführt. Die Grundlagen hierzu werde ich in diesem Abschnitt darstellen. Alle Analysen wurden auf Basis des von Rodríguez-Fdez u. a. [Rod+15] entwickelten Frameworks umgesetzt. Dabei werden die folgenden Signifikanzlevel unterschieden:

- Signifikant: $p \leq 0.05 \Rightarrow \star$
- Sehr signifikant: $p \leq 0.01 \Rightarrow \star\star$
- Hoch signifikant: $p \leq 0.001 \Rightarrow \star\star\star$

Für die vorgestellten und zu betrachtenden AUROC-Werte kann keine Normalenverteilung angenommen werden. Aus diesem Grund kann beispielsweise kein t-Test zum Testen von Signifikanzen verwendet werden. Anstatt dessen werden parameterfreie statistische Verfahren genutzt.

Ich verwende im Folgenden bei dem Vergleich von zwei Gruppen den *Wilcoxon Signed-Rank Test (WSR-Test)* [Wil45]. Dieser wird beispielsweise auch von Demšar [Dem06] und García u. a. [Gar+10] zur Evaluation von Klassifikatoren empfohlen. Der Test ist in der Lage die Nullhypothese „Es gibt keinen signifikanten Unterschied zwischen den beiden zu vergleichenden Gruppen“ zu testen.

Bei Vergleichen von Gruppengrößen $k > 2$ kann der WSR-Test allerdings nicht angewendet werden. Um Signifikanzen unter den Gruppen feststellen zu können, muss eine zweistufige statistische Analyse durchgeführt werden. Ich verwende zunächst einen *Friedman Aligned-Rank Test (FAR-Test)* [HL62; Gar+10], um zu testen, ob die Nullhypothese der gleichen Mediane in allen Gruppen abgelehnt werden kann. Im zweiten Schritt wird, wie auch von García u. a. [Gar+10] beschrieben, eine *Finner Post-hoc Analyse (FP-Analyse)* [Fin93] angewendet. Dabei wird durch eine Anpassung des p-Wertes zu p_{adj} , eine Fehlerakkumulierung von *Typ 1 Fehlern* (Falsch-Positiv-Fehler) bei gleichzeitigen mehrfachen Hypothesentests berücksichtigt. Dieses ist nötig, da die Wahrscheinlichkeit einen Typ 1 Fehler zu begehen bereits bei $k = 4$ Gruppen und einem Signifikanzlevel von $\alpha = 0.05$ bei 14 Prozent liegt. Bei $k = 4$ Gruppen werden pro betrachteter Gruppe beim Vergleich mit den anderen Gruppen drei Hypothesentests durchgeführt. Das ein Falsch-Positiv-Fehler in einem einzelnen Vergleich nicht auftritt, hat eine Wahrscheinlichkeit von $(1 - \alpha) = 0.95$. Die Wahrscheinlichkeit bei drei Vergleichen, einen Falsch-Positiv-Fehler nicht zu machen, ist $(1 - \alpha)^3 = 0.86$. Also besteht schon bei einer Gruppengröße von $k = 4$ eine Wahrscheinlichkeit von $1 - (1 - \alpha)^3 = 0.14$ die Nullhypothese fälschlicherweise abzulehnen,

obwohl keine Signifikanz vorliegt. Bei der im weiteren Verlauf dieses Kapitels betrachteten Gruppengröße von $k = 15$ ist diese Wahrscheinlichkeit bereits auf 51 Prozent angestiegen. Eine Fehlerakkumulierung muss also zwingend durch die Verwendung von adäquaten Verfahren verhindert werden.

6.2 Auswertung

In diesem Abschnitt werde ich die beschriebenen Klassifikationsmodelle (Bäume, Boosting, Nachbar, SVM), die auf unterschiedlichen Kombinationen der beschriebenen Merkmalsgruppen trainiert und optimiert wurden, durch die beschriebene Ermittlung der Güte von Klassifikatoren und den vorgestellten Testmethoden miteinander vergleichen. Die verwendeten Abkürzungen setzen sich aus den verwendeten Merkmalsgruppen (siehe Kapitel 4, Seite 39) zusammen. So werden beispielsweise bei *A-RSBZ* die Merkmalsgruppen A-R (Intensitäten der AUs), S (System-Events), B (Aufmerksamkeit) und Z (Dauer) verwendet. Analog dazu sind die Tabellen und Abbildungen im Folgenden aufgebaut.

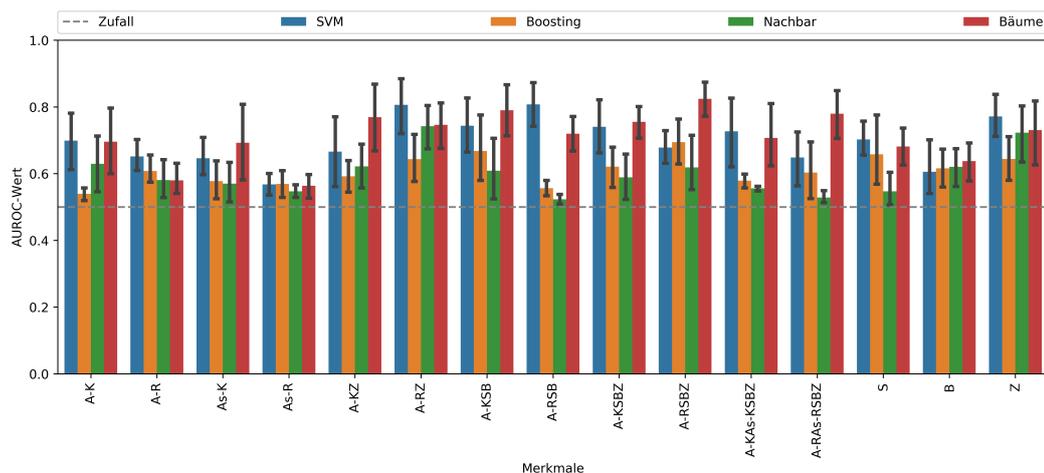


Abbildung 6.2.: Ergebnisse bei Variation der Klassifikationsketten und Merkmalsgruppen:

Es werden durchschnittliche AUROC-Werte und deren Standardabweichungen (schwarze Balken) der Kreuzvalidierungen der unterschiedlichen Klassifikationsmodelle bei Variation der Merkmalsgruppen dargestellt.

Die Ergebnisse der unterschiedlichen Klassifikationsmodelle, die auf der beschriebenen Datenbasis trainiert und optimiert wurden, sind in Tabelle 6.1 aufgeführt und in Abbildung 6.2 visualisiert. Die zugehörigen durchschnittlichen ROC Diagramme sind im Anhang A.2

Merkmale	SVM	Boost	Nachbar	Bäume
A-K	0.70 (± 0.11)	0.54 (± 0.03)	0.63 (± 0.10)	0.70 (± 0.12)
A-R	0.65 (± 0.06)	0.61 (± 0.05)	0.58 (± 0.07)	0.58 (± 0.06)
As-K	0.64 (± 0.07)	0.58 (± 0.08)	0.57 (± 0.08)	0.69 (± 0.15)
As-R	0.57 (± 0.04)	0.57 (± 0.05)	0.55 (± 0.02)	0.56 (± 0.05)
A-KZ	0.67 (± 0.13)	0.59 (± 0.06)	0.62 (± 0.08)	0.77 (± 0.13)
A-RZ	0.80 (± 0.11)	0.64 (± 0.09)	0.74 (± 0.08)	0.75 (± 0.09)
A-KSB	0.74 (± 0.10)	0.67 (± 0.12)	0.61 (± 0.11)	0.79 (± 0.10)
A-RSB	0.81 (± 0.08)	0.56 (± 0.03)	0.52 (± 0.02)	0.72 (± 0.07)
A-KSBZ	0.74 (± 0.10)	0.62 (± 0.08)	0.59 (± 0.09)	0.76 (± 0.06)
A-RSBZ	0.68 (± 0.06)	0.69 (± 0.09)	0.62 (± 0.10)	0.82 (± 0.06)
A-KAs-KSBZ	0.73 (± 0.12)	0.58 (± 0.03)	0.56 (± 0.01)	0.71 (± 0.12)
A-RAs-RSBZ	0.65 (± 0.10)	0.60 (± 0.11)	0.53 (± 0.02)	0.78 (± 0.09)
S	0.70 (± 0.06)	0.66 (± 0.13)	0.55 (± 0.07)	0.68 (± 0.07)
B	0.60 (± 0.11)	0.62 (± 0.07)	0.62 (± 0.07)	0.64 (± 0.07)
Z	0.77 (± 0.08)	0.64 (± 0.08)	0.72 (± 0.11)	0.73 (± 0.12)

Tabelle 6.1.: Übersicht aller ermittelten durchschnittlichen AUROC-Werte:

Es werden die durchschnittlichen AUROC-Werte und die Standardabweichungen der Kreuzvalidierungen der Klassifikationsmodelle bei verschiedenen Kombinationen der Merkmalsgruppen dargestellt [M (SD)]. Das beste Ergebnis ist grün, das schlechteste Ergebnis rot markiert.

(siehe Seite 146) dargestellt. Bei einer Betrachtung der gemittelten AUROC-Werte der einzelnen Kreuzvalidierungen je nach Klassifikatormodell und Merkmalsgruppenkombination, schneidet die Kombination A-RSBZ (Intensität der AUs + System-Events + Aufmerksamkeit + Dauer) in Kombination mit einer entscheidungsbaumbasierten Klassifikationskette am besten ab (siehe grün markierter Eintrag in Tabelle 6.1). Den schlechtesten Wert erzielt die Kombination A-RSB (Intensitäten der AUs + System-Events + Aufmerksamkeit) in Kombination mit einer nachbarschaftsbasierte Klassifikationskette (siehe rot markierter Eintrag in Tabelle 6.1).

6.2.1 Action Unit Auftreten vs. Action Unit Intensität

Um eine Aussage treffen zu können, ob die Betrachtung von Intensität oder Auftreten von AUs signifikant bessere Ergebnisse liefert (Hypothese 6.2), werden alle AUROC-Werte verwendet, die aus einem Trainings- und Testverfahren entstanden sind, das die jeweilige Merkmalsgruppe enthielt. Für die Betrachtung des AU Auftretens sind es die Merkmalsgruppen AU-K (Auftreten der AUs) und As-K (Summe der AU Auftritte) und für die Betrachtung der AU Intensitäten die Merkmalsgruppen AU-R (Intensitäten der AUs) und As-R (Summe der AU Intensitäten). Die Daten sind in Abbildung 6.3 visualisiert. Es können $n = 144$ Beispiele aus den einzelnen Kreuzvalidierungen herangezogen werden. Da die identische Datenbasis für beide Gruppen verwendet wird, sind diese abhängig voneinander.

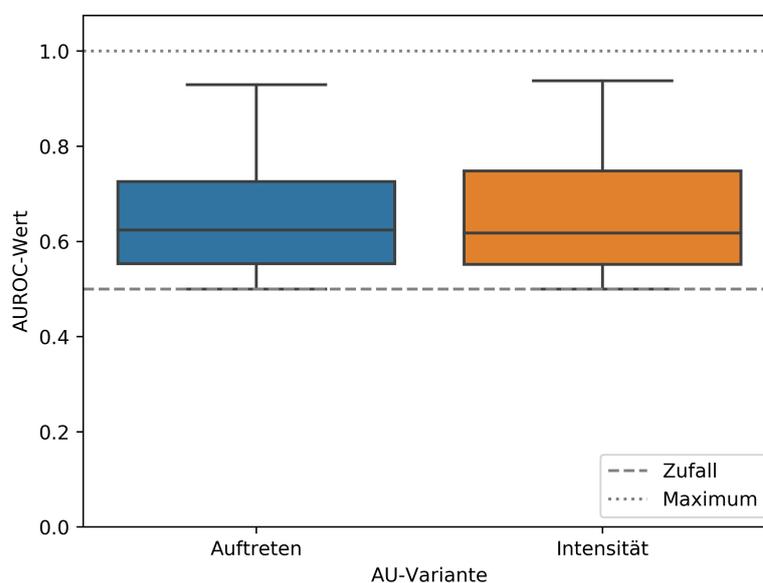


Abbildung 6.3.: Vergleich Action Unit Auftreten und Action Unit Intensität:

Es werden die durchschnittlichen AUROC-Werte bei der Betrachtung von Action Unit Auftreten und Action Unit Intensitäten dargestellt.

Die Nullhypothese des WSR-Tests (Es gibt keinen signifikanten Unterschied zwischen den beiden Gruppen.) kann nicht abgelehnt werden ($T = 5023.5, p = 0.81$). Die Ähnlichkeit der Gruppen wird auch bei der Betrachtung von Abbildung 6.3 deutlich und wird durch den WSR-Test bestätigt. Hypothese 6.2, dass sich basierend auf eine AU Auftrittserkennung bessere Ergebnisse erzielen lassen als auf einer AU Intensitätserkennung, kann also nicht bestätigt werden.

6.2.2 Vergleich der Klassifikationsmodelle

Um ein Klassifikatormodell und somit eine Verarbeitungskette bestimmen zu können, die die beste Leistung bringt, werden wiederum statistische Tests verwendet. Dabei werden die AUROC-Werte je nach zugrundeliegendem Modell der Klassifikatoren in Gruppen unterteilt. Es kann erneut nicht von einer Normalverteilung ausgegangen und die betrachteten Daten können wie zuvor als abhängig angenommen werden. Im Gegensatz zu der zuvor beschriebenen Analyse werden nun aber $k = 4$ Gruppen mit jeweils $n = 90$ AUROC-Werten je Gruppe betrachtet. Die Daten sind in Abbildung 6.4 dargestellt.

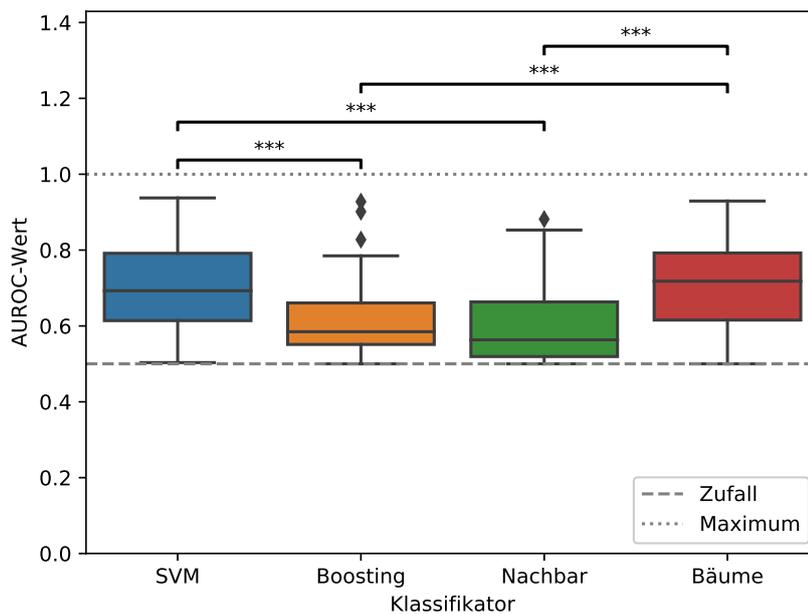


Abbildung 6.4.: Vergleich der Klassifikationsmodelle:

Es werden die durchschnittlichen AUROC-Werte bei der Variation der Klassifikationsmodelle und signifikante Unterschiede dargestellt.

Wie bereits dargestellt, ist bei Gruppengröße $k > 2$ der WSR-Test nicht anwendbar. Es wird das beschriebene zweistufige Testverfahren für größere Gruppengrößen angewendet. Zunächst wird mithilfe eines FAR-Test [HL62; Gar+10] gezeigt, dass die Nullhypothese der gleichen Mediane aller Gruppen mit hoher Signifikanz abgelehnt werden kann ($\chi^2 = 79.326, p < 0.00001$).

Im zweiten Schritt ergibt sich durch die FP-Analyse [Fin93] die in Tabelle 6.2 zu sehende Rangordnung der Klassifikationsmodelle. Baumbasierte Modelle erzielen dicht gefolgt von Modellen auf der Basis von Stützvektoren die besten Ergebnisse (je kleiner der Rang, desto

Klassifikator	Rang
Bäume	118.25
SVM	131.15
Boosting	231.31
Nachbar	241.29

Tabelle 6.2.: Rangordnung der Klassifikationsmodelle:

Es wird die durch eine Finner Post-hoc Analyse ermittelte statistische Rangordnung der Klassifikationsmodelle dargestellt.

besser ist der Klassifikator). Einen deutlichen Abstand haben die Modelle basierend auf Nachbarschaftsbetrachtung und Boosting.

Die FP-Analyse lässt auf hoch signifikante Unterschiede zwischen den folgenden Paaren schließen:

- Baum vs. Nachbar ($Z = 7.93, p_{adj} < 0.00001$)
- Baum vs. Boosting ($Z = 7.29, p_{adj} < 0.00001$)
- SVM vs. Nachbar ($Z = 7.10, p_{adj} < 0.00001$)
- SVM vs. Boosting ($Z = 6.46, p_{adj} < 0.00001$).

Die Signifikanzen sind auch in Abbildung 6.4 dargestellt. Signifikante Unterschiede zwischen Modellen auf Basis von Nachbarschaft und Boosting ($Z = 0.64, p_{adj} = 0.52$) und zwischen SVM- und baumbasierten Modellen ($Z = 0.83, p_{adj} = 0.46$) können nicht festgestellt werden. Somit sind SVM-basierte und baumbasierte Modelle signifikant besser als die anderen beiden getesteten Klassifikationsansätze. Welches dieser beiden Verfahren aber das beste Modell ist, kann nicht gezeigt werden. Die Rangordnung lässt zwar vermuten, dass baumbasierte Verfahren in der vorgestellten Anwendung die besten Ergebnisse liefern, eine Signifikanz ist jedoch auf der betrachteten Datenbasis nicht erkennbar.

6.2.3 Kombination verschiedener Merkmalsgruppen

Zum Vergleichen unterschiedlicher Merkmalsgruppenkombinationen werden alle Klassifikationsmodelle mit der selben Merkmalsgruppenkombination zusammengefasst. Pro betrachteter Gruppe stehen durch die einzelnen Faltungen $n = 24$ AUROC-Werte zur Verfügung und es werden $k = 15$ Gruppen getestet. Die Daten sind ohne berechnete Signifikanzen in Abbildung 6.5 visualisiert.

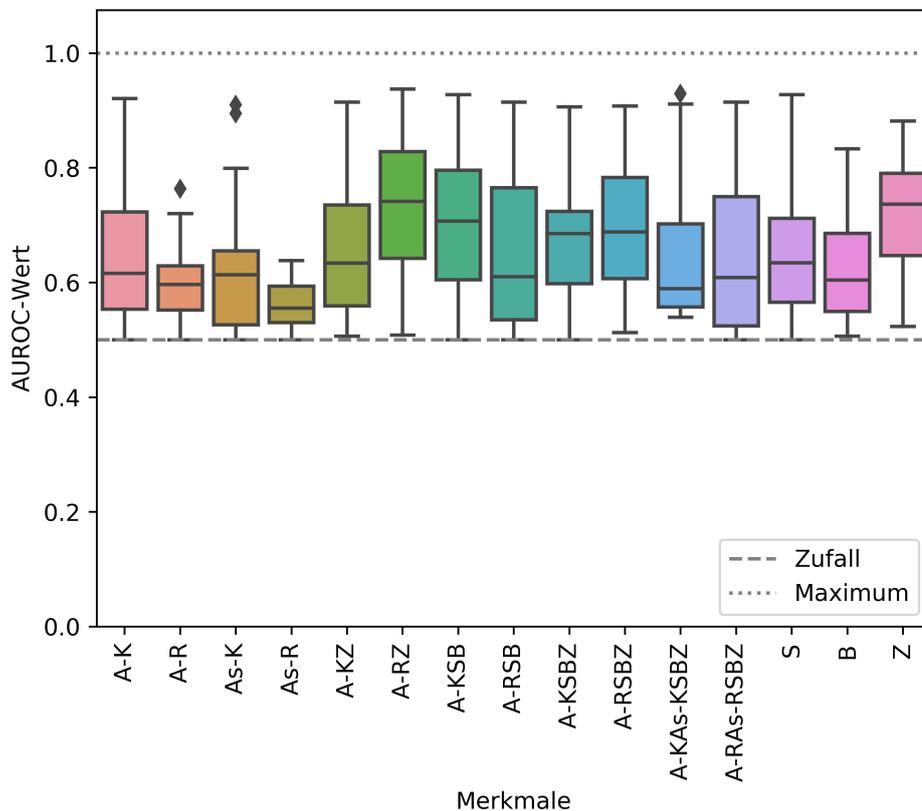


Abbildung 6.5.: Vergleich der Merkmalsgruppenkombinationen:

Es werden die AUROC-Werte bei Variation der Kombination unterschiedlicher Merkmalsgruppen dargestellt.

Durch den durchgeführten FAR-Test kann eine hohe Signifikanz für einen Unterschied mindestens zweier der Gruppen festgestellt werden ($\chi^2 = 65.48, p < 0.00001$). Somit lässt sich erneut die Durchführung einer Post-Hoc Analyse (wiederum FP-Analyse) motivieren.

Die ermittelten statistischen Ränge der einzelnen Merkmalsgruppen sind in Tabelle 6.3 zu sehen. Die Merkmalsgruppenkombination A-RZ (Intensitäten der AUs + Dauer) schneidet am besten ab. Sie wird gefolgt von der alleinigen Verwendung der Dauer (Merkmalsgruppe Z) und der multimodalen Merkmalsgruppenkombination A-RSBZ (Intensitäten der AUs + System-Events + Aufmerksamkeit + Dauer). Die schlechteste Merkmalsgruppe ist nach der Rangordnung die Gruppe As-R (Summe der AU Intensitäten).

Alle ermittelten signifikanten Unterschiede sind in Tabelle 6.4 zu sehen (Signifikanzlevel wie beschrieben: $p \leq 0.05 \Rightarrow *$, $p \leq 0.01 \Rightarrow **$, $p \leq 0.001 \Rightarrow ***$). Ich beschreibe

Merkmale	Rang
A-RZ	99.85
Z	118.77
A-RSBZ	127.65
A-KSB	128.38
A-KSBZ	152.23
A-KZ	174.69
A-RSB	187.94
A-KAs-KSBZ	190.88
S	194.38
A-K	197.50
A-RAs-RSBZ	200.25
B	209.75
As-K	223.85
A-R	227.50
As-R	273.90

Tabelle 6.3.: Rangordnung der Merkmalsgruppenkombinationen:

Es wird die durch eine Finner Post-hoc Analyse ermittelte statistische Rangordnung der Merkmalsgruppenkombinationen dargestellt.

für eine anschaulichere Analyse nur die Signifikanzen einzelner und herausstechender Merkmalsgruppenkombinationen (siehe Abbildung 6.6).

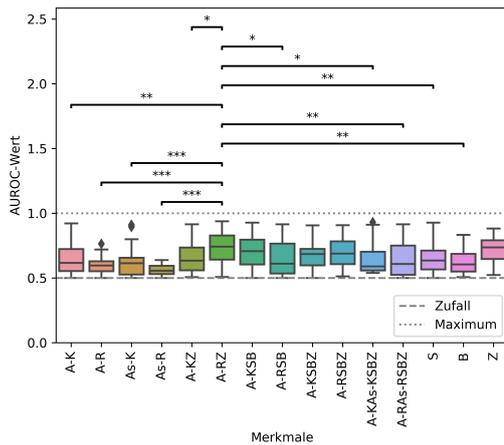
Abbildung 6.6a stellt alle Signifikanzen der Merkmalsgruppenkombination A-RZ (Intensitäten der AUs + Dauer) dar. Diese hat den höchsten Rang in der statistischen Auswertung. Es lassen sich zu 10 der anderen 14 Merkmalsgruppenkombination signifikante Unterschiede feststellen. Einzig zu der Merkmalsgruppe Z (Dauer) ($Z = 0.63, p_{adj} = 0.61$), der Merkmalsgruppenkombination A-KSB (Auftreten der AUs + System-Events + Aufmerksamkeit) ($Z = 0.95, p_{adj} = 0.46$), der Merkmalsgruppenkombination A-KSBZ (Auftreten der AUs + System-Events + Aufmerksamkeit + Dauer) ($Z = 1.74, p_{adj} = 0.16$) und der Merkmalsgruppenkombination A-RSBZ (Intensitäten der AUs + System-Events + Aufmerksamkeit + Dauer) ($Z = 0.93, p_{adj} = 0.47$) ist keine Signifikanz erkennbar und die Nullhypothese einer gleichen Mediane lässt sich nicht ablehnen.

Auf dem zweiten Platz der Rangordnung ist die reine Betrachtung der Dauer (Z) als Merkmal zu finden. Die Signifikanzen dieser Merkmalsgruppe sind in Abbildung 6.6b zu sehen. Hierbei konnten zu 8 der anderen 14 Merkmalsgruppen signifikante Unterschiede gefunden werden.

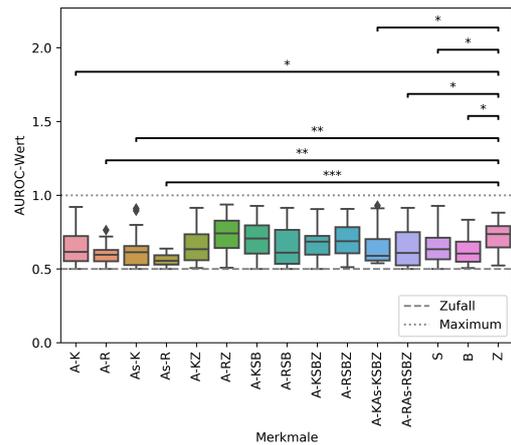
	A-K	A-R	As-K	As-R	A-KZ	A-RZ	A-KSB	A-RSB	A-KSBZ	A-RSBZ	A-KAs-KSBZ	A-RAs-RSBZ	S	B	Z
A-K				*		**									*
A-R						***	**		*	**					***
As-K						***	**		*	**					**
As-R	*				**	***	***	*	***	***	*	*	*	*	***
A-KZ				**											
A-RZ	**	***	***	***	*			*			*	**	**	**	
A-KSB		**	**	***										*	
A-RSB			*	*		*						*			
A-KSBZ		*	*	***											
A-RSBZ		**	**	***							*	*	*	*	
A-KAs-KSBZ				*		*									*
A-RAs-RSBZ				*		**	*			*	*				*
S				*		**									*
B						**	*			*					*
Z	*	**	**	***							*	*	*	*	*

Tabelle 6.4.: Signifikanzen aller Merkmalsgruppenkombinationen:

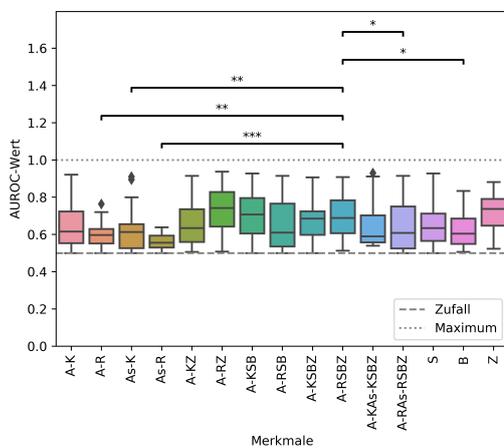
Es werden alle ermittelten signifikanten Unterschiede zwischen einzelnen Merkmalsgruppenkombinationen dargestellt (Signifikanzlevel $p \leq 0.05 \Rightarrow *$, $p \leq 0.01 \Rightarrow **$ und $p \leq 0.001 \Rightarrow ***$).



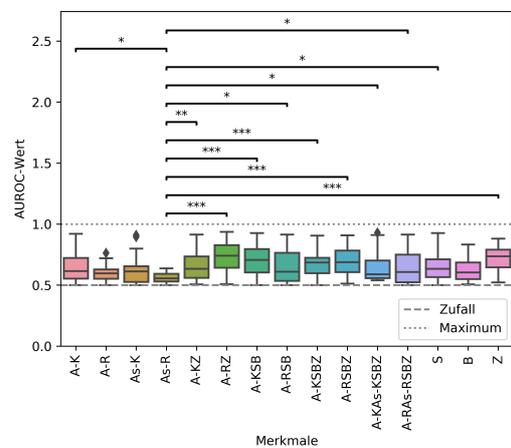
(a) Durchschnittliche AUROC-Werte und Signifikanzen auf Basis der Kombination der Merkmalsgruppen A-R und Z



(b) Durchschnittliche AUROC-Werte und Signifikanzen auf Basis der Merkmalsgruppe Z



(c) Durchschnittliche AUROC-Werte und Signifikanzen auf Basis der Kombination der Merkmalsgruppen A-R, S, B und Z



(d) Durchschnittliche AUROC-Werte und Signifikanzen auf Basis der Merkmalsgruppe As-R

Abbildung 6.6.: Vergleich und Signifikanzen von Merkmalsgruppenkombinationen:

Es werden die durchschnittliche AUROC-Werte bei der Variation der Kombination unterschiedlicher Merkmalsgruppen und einzelne Signifikanzen dargestellt.

Bei der darauf folgenden Merkmalsgruppenkombination A-RSBZ (Intensitäten der AUs + System-Events + Aufmerksamkeit + Dauer) können nur noch zu 5 der anderen 14 Gruppen Signifikanzen festgestellt werden (siehe Abbildung 6.6c). Wie in Tabelle 6.1 (siehe Seite 88) zu sehen ist, erzielt diese Merkmalsgruppenkombination in Verbindung mit einer baumbasierten Klassifikationskette den besten AUROC-Wert von 0.82 (SD= ± 0.06). Dennoch kann bei der gemeinsamen Betrachtung aller Klassifikationsmodelle nur das drittbeste

Ergebnis erzielt werden, da die Leistung der drei weiteren Klassifikationsverfahren bei der Verwendung dieser Merkmalsgruppenkombination gering ist (siehe Tabelle 6.1, Seite 88).

Die Signifikanzen der in der Rangordnung schlechtesten Kombination von Merkmalsgruppen sind in Abbildung 6.6d zu sehen. Zu 11 der anderen 14 Merkmalsgruppen konnten signifikante Unterschiede festgestellt werden. Nur im Vergleich zur Verwendung der Merkmalsgruppe A-R (Intensitäten der AUs) ($Z = 1.54, p_{adj} = 0.22$), der Merkmalsgruppe As-K (Summe der AU Auftreten) ($Z = 1.67, p_{adj} = 0.18$) und der Merkmalsgruppe B (Aufmerksamkeit) ($Z = 2.14, p_{adj} = 0.08$) konnte die Nullhypothese der gleichen Mediane nicht verworfen werden und es wurden ähnlich schlechte Ergebnisse erzielt.

Darüber hinaus kann auch durch den Merkmalsgruppenvergleich der zuvor gezeigte geringe Unterschied zwischen der Nutzung von AU Intensitäten und AU Auftreten bei sonst gleichen weiteren verwendeten Merkmalsgruppen festgestellt werden. Nur bei dem Vergleich zwischen der Kombination A-KZ (Auftreten der AUs + Dauer) und der Kombination A-RZ (Intensitäten der AUs + Dauer) tritt ein signifikanter Unterschied auf ($Z = 2.49, p_{adj} = 0.04$). Bei allen anderen Vergleichen zwischen denselben Merkmalsgruppen bei einer alleinigen Variation der AU Erkennungsmethode treten keine signifikanten Unterschiede auf.

6.2.4 Qualitative Fehleranalyse

Die zuvor dargestellten AUROC-Werte lassen in den meisten Fällen auf eine Klassifikatorleistung im befriedigenden Bereich (0.70–0.80) schließen. Dieser nicht optimale Wert kann einerseits dadurch erklärt werden, dass dasselbe menschliche Verhalten in beiden Klassen gezeigt wird und sich somit die definierten Merkmalsgruppen nicht für eine Erkennung von Aufgabenschritten mit schwerwiegenden Störungen eignen. Andererseits kann es aber auch sein, dass die gelernten Klassifikationsketten die einzelnen Klassen trennen können, aber einzelne Prozessschritte, die keine manuell eingefügten Prozessstörungen enthalten, dennoch der Klasse der Störungen logisch richtig zugeordnet werden, da in diesen Schritten tatsächlich auch Prozessstörungen auftreten. Die Klasseneinteilung wäre somit nicht optimal. Das würde bedeuten, dass der durch die nutzerorientierte Entwicklung erstellte Prozess nicht optimal war und das Verfahren in der Lage wäre, weitere Prozessstörungen aufzuzeigen.

Dieser Abschnitt geht der Frage nach, welche Prozessschritte der falschen Klasse zugeordnet werden und warum dieses der Fall ist. Dafür analysiere ich qualitativ die falschen Klassifikationen der gelernten Verarbeitungsketten, um eventuelle Regelmäßigkeiten in den falschen Entscheidungen der Klassifikatoren zu finden.

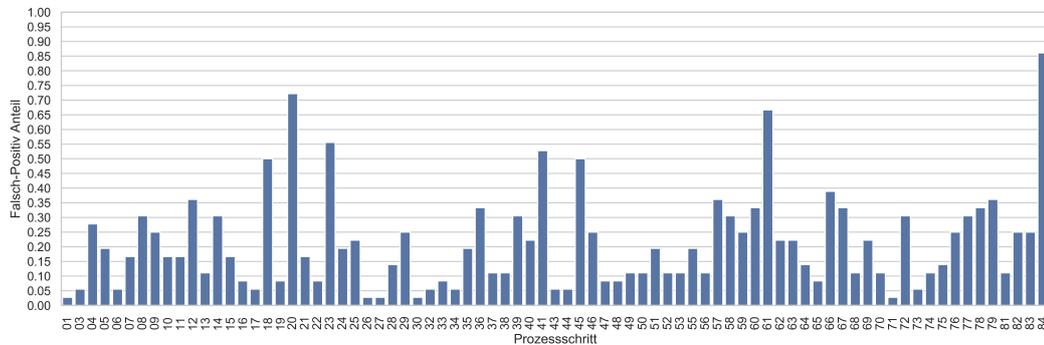


Abbildung 6.7.: Häufigkeit einer falschen Klassifikation je Prozessschritt:

In einem normierten Histogramm wird die Häufigkeit einer falschen Klassifikation je Prozessschritt bei Anwendung der drei besten Merkmalsgruppenkombination in Verbindung mit den beiden besten Klassifikationsketten dargestellt.

Um eine Übersicht über diese Prozessschritte zu erstellen, wurden die Ergebnisse der Kreuzvalidierungen der drei besten Merkmalsgruppenkombinationen (A-RZ [Intensitäten der AUs + Dauer], Z [Dauer], A-RSBZ [Intensitäten der AUs + System-Events + Aufmerksamkeit + Dauer]) jeweils in Kombination mit den beiden besten Klassifikationsmodellen (SVM, Bäume) genauer betrachtet und es wurde ein normiertes Histogramm über sämtliche Prozessschritte der falschen Klassifikationen erstellt (siehe Abbildung 6.7).

Sieben Aufgabenschritte wurden in mehr als der Hälfte aller Klassifikationen falsch klassifiziert. Dieses sind in absteigender Reihenfolge Schritt 84 (Fehlerrate 0.86), Schritt 20 (Fehlerrate 0.72), Schritt 61 (Fehlerrate 0.67), Schritt 23 (Fehlerrate 0.56), Schritt 41 (Fehlerrate 0.53), Schritt 45 (Fehlerrate 0.5) und Schritt 18 (Fehlerrate 0.5).

Schritt 84 ist dabei einer der zuvor modellierten schwerwiegenden Prozessstörungen. Die modellierten leichten Prozessstörungen (Schritt 11, Schritt 60, Schritt 62, Schritt 63) weisen keine erhöhte Fehlerrate auf. Ein Auftreten von einem ähnlichen Kopfverhalten bei leichten Prozessstörungen im Vergleich mit Prozessschritten mit schwerwiegenden Störungen lässt sich nicht erkennen. Im Folgenden werde ich diese häufigsten Fehlerquellen diskutieren und analysieren:

Schritt 84

In diesem Schritt erfolgt die Sichtkontrolle der geleisteten Arbeit durch die Versuchsperson (siehe Abbildung 5.6, Seite 68). Durch die Komplexität und die Masse der ausgeführten Schritte vor dieser Sichtkontrolle wurde angenommen, dass die Versuchsperson überfordert ist und dadurch andere Verhaltensweisen zeigt. Die sehr große Anzahl an falschen Klassifikationen dieses Schrittes lässt darauf schließen, dass sich das gezeigte Verhalten deutlich von den sonstigen definierten schwerwiegenden Prozessstörungen unterscheidet. Eine Sichtung des Videomaterials dieses Schrittes deckt auf, dass besonders die gezeigten Kopfausdrücke (zumeist Lächeln und Feixen) sich deutlich von anderen schwerwiegenden Prozessstörungen unterscheiden (u. a. Zusammenziehen der Augenbraun). Ein Grund hierfür könnte sein, dass die Person die Anweisung sehr gut versteht, aber die Umsetzung für unmöglich hält. Auch die Art der Aufgabe ist eine andere als die zuvor bearbeiteten Schritte. In diesem Fall ist die Aufgabenstellung sehr offen gehalten und es gibt keine Möglichkeit zur direkten Überprüfung der korrekten Durchführung der Sichtkontrolle innerhalb des Systems. Außerdem arbeiten die Versuchspersonen zu diesem Zeitpunkt schon annähernd 60 Minuten an dem Motor. Dadurch kann die Motivation einer Kontrolle gering sein, da sie in erste Linie die Studie beenden möchten. Die Qualität des gebauten Produktes wurde darüber hinaus nicht mit der Höhe der Aufwandsentschädigung verbunden. Es kann also dazu kommen, dass eine gute Qualitätsprüfung ignoriert wird, da keine persönlichen Konsequenzen aus einer schlechten Aufgabenerfüllung hervorgehen. Die Wahrscheinlichkeit ist somit hoch, dass sich die Merkmale so stark von den anderen definierten schwerwiegen-



Abbildung 6.8.: 20. Schritt des Montageprozesses:

Es ist die durch das System gegebene Beschreibung des zu erfüllenden 20. Schrittes des Montageprozesses zu sehen.

den Prozessstörungen unterscheiden und der negativen Klasse ähnlicher sind. Eine Erkennung eines solchen Schrittes wäre aber dennoch sinnvoll und wünschenswert.

Schritt 20

Dieser Schritt (siehe Abbildung 6.8) ist keiner der beschriebenen schwerwiegenden Prozessstörungen. Dennoch ist in der Modellierung ein logischer Fehler aufgetreten, der in der Systementwicklung nicht aufgefallen ist. Der Schritt dient dem Fixieren einer Mutter an einer zuvor eingesetzten Schraube. Die zu verwendende Mutter ist jedoch nicht in der Stückliste angegeben und auch auf dem Bild ist diese nicht zu sehen, da sie sich in dem Werkzeug befindet. Es ist also wahrscheinlich, dass sich das Verhalten der Versuchspersonen in diesem Schritt sehr den Verhaltensweisen in den manuell hinzugefügten schwerwiegenden Prozessstörungen ähnelt. Somit ist es auch nicht verwunderlich, dass dieser Prozessschritt häufig als schwerwiegende Störung erkannt wird.

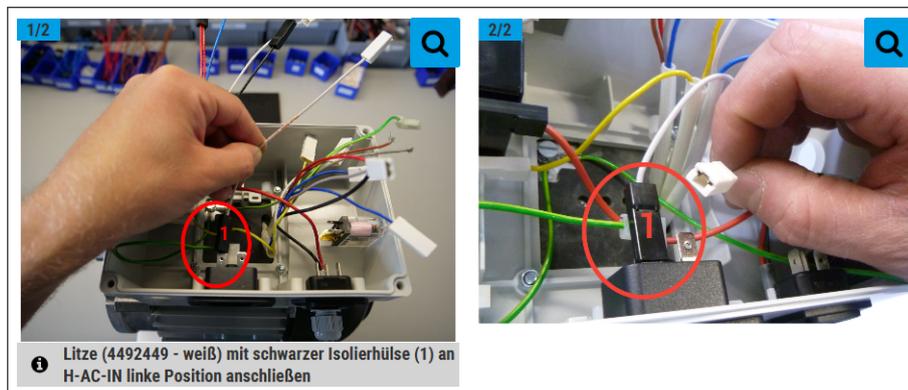


Abbildung 6.9.: 61. Schritt des Montageprozesses:

Es ist die durch das System gegebene Beschreibung des zu erfüllenden 61. Schrittes des Montageprozesses zu sehen.

Schritt 61

Die Darstellung des Aufgabenschrittes 61 ist in Abbildung 6.9 zu sehen. Es muss eine Litze mit einer bestimmten Seite an einen Stecker angeschlossen werden. Durch das gezeigte Bild wird nicht deutlich, dass neues Material verwendet werden muss. Es sieht vielmehr so aus, als ob die Litze bereits an dem Eingang befestigt ist (siehe linke Bild in Abbildung 6.9). Somit wurde eine Beeinträchtigung im Systemablauf von den Klassifikatoren erkannt, die durchaus berechtigt ist.

Schritt 23 / Schritt 18

Schritt 23 (siehe Abbildung 5.4b, Seite 67) und Schritt 18 (siehe Abbildung 6.10) sind

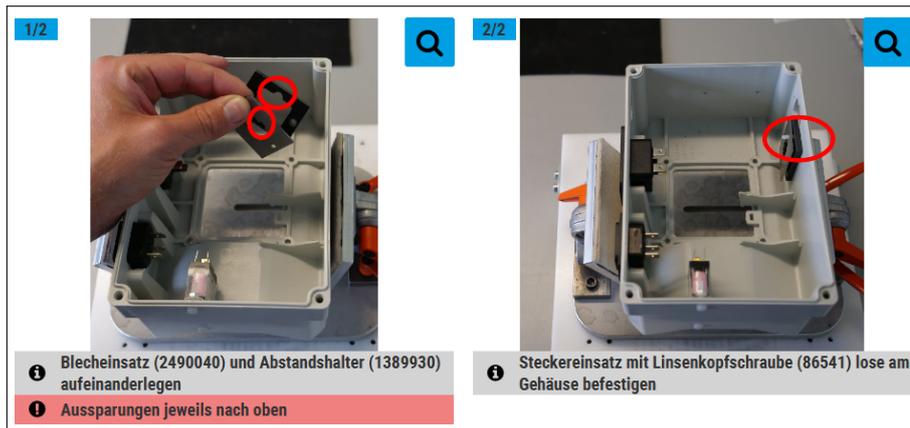


Abbildung 6.10.: 18. Schritt des Montageprozesses:

Es ist die durch das System gegebene Beschreibung des zu erfüllenden 18. Schrittes des Montageprozesses zu sehen.

sich sehr ähnlich. Bei einer Analyse der Schritte fällt auf, dass von der Versuchsperson mehrere Handlungen gefordert sind. Die Vermutung liegt nahe, dass diese Inkonsistenz im Vergleich zu den anderen Prozessschritten den Nutzer verwirren. Außerdem werden diese Schritte in der Durchführung länger dauern als die übrigen Schritte, in denen nur eine Aufgabe erfüllt werden muss. Somit kann auch dieser Schritt den gelernten schwerwiegenden Prozessstörungen ähnlicher sein.

Schritt 45

In diesem Schritt (siehe Abbildung 6.11) wird der montierte Kasten auf den restlichen Motor aufgesetzt. Diese Aufgabe ist nicht einfach, da dabei vier variable Schrauben in dazugehörige Aussparungen im Kasten gesteckt werden müssen und gleichzeitig das

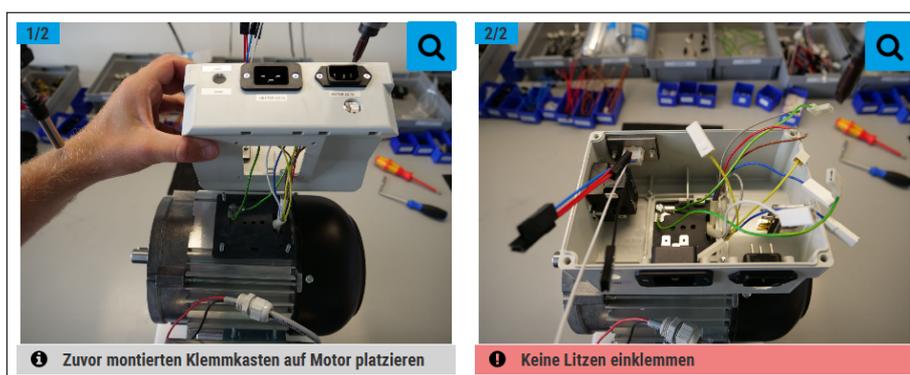


Abbildung 6.11.: 45. Schritt des Montageprozesses:

Es ist die durch das System gegebene Beschreibung des zu erfüllenden 45. Schrittes des Montageprozesses zu sehen.

Einklemmen von Litzen verhindert werden muss. Trotz der korrekten Anweisung scheint der Nutzer in diesem Schritt vor Problemen zu stehen.

6.3 Zusammenfassung und Diskussion

In diesem Kapitel habe ich dargestellt, ob und wie Prozessschritte mit manuell definierten Prozessstörungen durch einen überwachten Klassifikationsansatz erkannt werden können. Dabei habe ich gruppenbasierte Kreuzvalidierungen verwendet, um die Güte verschiedener Klassifikationsketten miteinander zu vergleichen.

Bei der Betrachtung des durchschnittlichen AUROC-Wertes über die Faltungen der Kreuzvalidierung schneidet eine multimodale Merkmalsgruppenkombination von AU Intensitäten, System-Events, Aufmerksamkeit und Dauer in Kombination mit Entscheidungsbäumen am besten ab. Am schlechtesten schneidet die multimodale Kombination von AU Intensitäten, System-Events und Aufmerksamkeit im Zusammenspiel mit nachbarschaftsbasierten Klassifikationsverfahren ab. Die erreichten AUROC-Werte befinden sich in den meisten Fällen im befriedigenden Bereich. Nur selten können Werte im guten Bereich erzielt werden. Hypothese 6.1 (Erkennbarkeit von Schritten mit schwerwiegenden Prozessstörungen) kann aber bestätigt werden. Es stellt sich dennoch die Frage, ob überwachte Verfahren die richtige Wahl für diese Art von Problem ist.

Es konnten keine signifikanten Unterschiede zwischen AU Intensitätenerkennung im Vergleich zu AU Auftrittserkennung gefunden werden. Ich kann also keine Empfehlung geben, welche der Merkmale am besten verwendet werden sollte. Somit kann Hypothese 6.2, dass basierend auf eine AU Auftrittserkennung bessere Ergebnisse möglich sind als bei der Verwendung einer AU Intensitätserkennung nicht bestätigt werden. In Abbildung 6.3 (siehe Seite 89) ist zu sehen, dass beide Kategorien stark streuen. Dieses Problem könnte durch eine größere Datenbasis gelöst werden. Weitere Aufnahmen sind aber im Rahmen dieser Dissertation nicht möglich.

Es konnte auch kein signifikanter Unterschied zwischen stützvektor- und baumbasierten Verarbeitungsketten nachgewiesen werden. Dennoch lässt sich feststellen, dass beide Verfahren signifikant besser abschneiden als nachbarschafts- und boostingbasierte Verfahren und somit diesen vorgezogen werden sollten. Auch hier könnte durch eine größere Datenmenge die Varianz verringert werden, um somit evtl. einen signifikanten Unterschied feststellen zu können.

Auch eine statistische Analyse unterschiedlicher Merkmalsgruppenkombinationen konnte keine eindeutig beste Kombination ermitteln. Beim Vergleich der einzelnen Merkmalsgruppen treten etliche Signifikanzen auf, klar überlegen gegenüber sämtlichen anderen Gruppen ist jedoch keine Kombination. Es lassen sich aber dennoch Schlussfolgerungen aus der Merkmalsgruppenanalyse ableiten. So sollte auf die alleinige Nutzung der Merkmalsgruppen As-R (Summe der AU Intensitäten), As-K (Summe der AU Auftritte), A-R (Intensitäten der AUs) und B (Aufmerksamkeit) verzichtet werden. Diese liefern die schlechtesten Ergebnisse in der Analyse. Somit lässt sich auch Hypothese 6.3 (Zusammenfassung der AUs zu einem einzelnen Merkmal reicht zur Erkennung von Störungen nicht aus) bestätigen, da die Merkmalsgruppen As-K und As-R mit die schlechtesten Ergebnisse liefern.

Bei alleiniger Betrachtung des Rangs ist die scheinbar beste Merkmalsgruppenkombination die Kombination von A-R (Intensitäten der AUs) und Z (Dauer). Es können aber keine statistischen Signifikanzen zu den Kombinationen A-KSB (Auftreten der AUs + System-Events + Aufmerksamkeit), A-KSBZ (Auftreten der AUs + System-Events + Aufmerksamkeit + Dauer), A-RSBZ (Intensitäten der AUs + System-Events + Aufmerksamkeit + Dauer) und der alleinigen Verwendung von Merkmalsgruppe Z (Dauer) festgestellt werden. Daraus folgt, dass mit dem vorhandenen Datensatz keine Aussage zu der besten Merkmalsgruppenkombination abgeleitet werden kann. Da die Kombination A-R (Intensitäten der AUs) und Z (Dauer) den höchsten Rang erzielt und mit die kleinste Dimensionalität aufweist, empfehle ich anhand dieses Datensatzes eine Verwendung dieser Kombination. Streng statistisch gesehen kann aber keine Empfehlung gegeben werden, da die Leistungen statistisch nicht differenziert werden können. Es lässt sich feststellen, dass im Allgemeinen eine Kombination mehrere Merkmalsgruppen bessere Ergebnisse liefert als die Benutzung einer einzelnen Gruppe. Nahezu alle Klassifikationssysteme mit den besten Resultaten basieren auf einer multimodalen Kombination von Merkmalsgruppen. Allein die Nutzung der Dauer sticht als einzelnes Merkmal heraus. Somit kann Hypothese 6.4, dass eine multimodale Betrachtung der Merkmalsgruppen die besten Ergebnisse liefert, dennoch bestätigt werden.

Ein Auftreten von vermehrt falschen Klassifikationen der auf schwerwiegenden Störungen trainierten Verarbeitungsketten bei der Anwendung auf leichte Prozessstörungen konnte nicht gezeigt werden. Es lässt sich vermuten, dass die leichten Störungen zu gering waren und die Versuchsperson nicht beeinflusst haben. Die leicht angepassten bzw. entfernten Informationen scheinen während des jeweiligen Schrittes nicht benötigt zu werden und können somit auch in der zukünftigen Anwendung des praktischen Systems weggelassen

werden. Somit stellen die modellierten leichten Störungen vielmehr eine Optimierung der Beschreibung auf die wesentlichen Information dar und führen nicht zu der angestrebten Verunsicherung der Versuchspersonen.

Abschließend konnte ich durch eine qualitative Analyse erklären, warum die gelernten Klassifikationsketten eine nicht optimale Leistung erbringen und warum somit relativ viele falsche Klassifikationen vorkommen. Dazu wurden die gemachten Fehler der besten Klassifikationsketten analysiert. Es konnte festgestellt werden, dass einige Prozessschritte nahezu in jeder betrachteten Klassifikatorkonfiguration der falschen Klasse zugeordnet werden. Bei der detaillierten Analyse dieser Schritte wurde deutlich, dass die gemachte Einteilung zwischen schwerwiegenden Störungen und allen anderen Schritten nicht haltbar ist. Zum einen wurde ein modellierter schwerwiegender Fehler in den meisten Fällen nicht als Fehler erkannt, zum anderen kam es in einigen der angenommenen störungsfreien Prozessschritten fast immer zur Erkennung einer schwerwiegenden Prozessstörung.

Es ist aber festzuhalten, dass die Fehler der Klassifikation qualitativ erklärbar und die Klassifikationen durchaus sinnvoll sind. Dieses lässt darauf schließen, dass die gewählten Merkmalsgruppen zur Erkennung von Störungen sinnvoll sind. Die gewählte Klasseneinteilung scheint jedoch nicht optimal zu sein. Aus diesem Grund werde ich im folgenden Kapitel ein Verfahren vorstellen, das auf keiner vorgegebener Klasseneinteilung beruht.

Erkennung von Prozessstörungen durch unüberwachte Verfahren

“ *The goal is to turn data into information, and
information into insight.*

– Carly Fiorina

Im vorherigen Kapitel habe ich einen überwachten Klassifikationsansatz zur Erkennung von Prozessschritten mit schwerwiegenden Störungen vorgestellt. Ich habe gezeigt, dass die Erkennung in angemessener Weise funktioniert, es aber dennoch zu falschen Klassifikationen während der Kreuzvalidierung kommt. Diese Fehler sind durch eine qualitative Analyse erklärbar und das Ergebnis der jeweiligen Klassifikation ist durchaus sinnvoll. So kann es sein, dass beispielsweise Fehlerstufen falsch abgeschätzt wurden. Darüber hinaus können im verwendeten und als optimal angesehenen Prozess bereits Störungen vorhanden sein, die nicht manuell hinzugefügt und somit im überwachten Verfahren nicht den Störungen, sondern der störungsfreien Klasse zugeordnet wurden. Dieses lässt darauf schließen, dass die zugrundeliegende Klasseneinteilung nicht optimal war. Eine saubere Einteilung der Klassen wäre nur mit sehr viel Annotationsaufwand durchführbar und kann scheinbar nicht durch einen nutzerorientierten Entwicklungsprozess erreicht werden.

In diesem Kapitel werde ich Methoden vorstellen, die keine zuvor definierten Klassen-zugehörigkeiten lernen, sondern unüberwacht ein *normales Nutzungsverhalten* während eines Prozessschrittes lernen und Abweichungen zu diesem Verhalten erkennen können. Dabei kommt erneut der zuvor genutzte Korpus aus 13 Probanden zur Anwendung. Durch eine detaillierte Analyse der aufgenommenen Daten wurde festgestellt, dass Menschen

ein ähnliches Verhalten während der Lösung von Prozessschritten mit einer ähnlichen Komplexität zeigen und bei Störungen des Prozesses Abweichungen zu diesem Verhalten auftreten. Besonders der Gesichtsausdruck ist bei störungsfreien Prozessen neutral und ändert sich kaum. Erst bei Schwierigkeiten kommt es zu einer anderen Mimik der jeweiligen Versuchsperson.

Es werden folgende Hypothesen betrachtet:

- Hypothese 7.1** Durch unüberwachte Verfahren können Prozessstörungen detektiert werden.
- Hypothese 7.2** Die Betrachtung von multimodaler Merkmalsgruppenkombinationen hat einen Mehrwert gegenüber der reinen zeitlichen Analyse.
- Hypothese 7.3** Es lässt sich ein Werkzeug entwickeln, das Prozessentwickler bei der Optimierung von Prozessabläufen unterstützt.

Wie bereits im Kapitel zuvor, werde ich auch im Folgenden unterschiedliche Merkmalsgruppen miteinander kombinieren und die auf Basis dieser Gruppenkombinationen gelernten Modelle miteinander vergleichen.

7.1 Systemaufbau

Das normale Verhalten der Versuchspersonen während einzelner Prozessschritte bilde ich durch ein Gauß-Modell ab. Da die Merkmalsgruppen bereits zeitlich normiert sind, können einzelne Schritte unterschiedlicher Durchführungsdauer direkt miteinander in Verbindung gebracht werden. Es ist zu beachten, dass im Folgenden je nach Merkmalsgruppenkombination multi- und eindimensionale Gauß-Modelle betrachtet werden.

Wie in Abbildung 7.1 zu sehen ist, werden alle Prozessschritte aller Personen des vorhandenen Datensatzes verwendet, um das normale Verhalten der Versuchspersonen durch ein Modell auf Basis einer Gauß-Verteilung zu schätzen. Zur Erstellung des Gauß-Modells wird pro Merkmalsgruppe bzw. pro Merkmalsgruppenkombination zunächst das durchschnittliche Vorkommen berechnet. Außerdem wird mithilfe der von Rousseeuw [Rou84] vorgestellten und durch Rousseeuw und van Driessen [Rv99] optimierten *Minimum Covariance Determinant (MCD)* eine robuste Kovarianzapproximation durchgeführt. Im Gegensatz zu einer empirischen Kovarianzberechnung werden durch die MCD Berechnung Ausreißer durch eine geringe Gewichtung berücksichtigt.

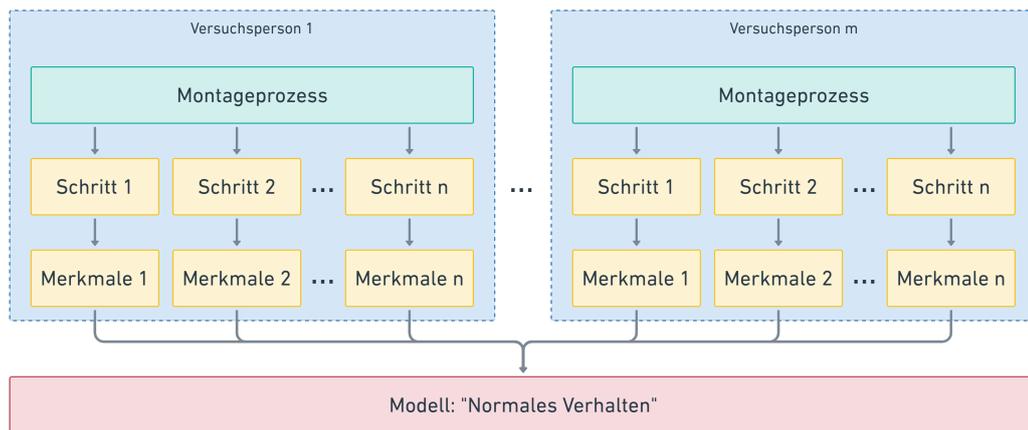


Abbildung 7.1.: Verfahren zum Erlernen eines Modells „normalen Verhaltens“:
 Für jeden Prozessschritt des Montageprozesses werden für jede Versuchsperson Merkmale extrahiert und in ein gemeinsames Modell überführt.

Die Bewertung einzelner Prozessschritte ist schematisch in Abbildung 7.2 dargestellt. Zur Abstandsberechnung wird die Mahalanobis-Distanz [Mah36] verwendet. Für eine Bewertung von *Schritt X* (siehe Abbildung 7.2) werden die relevanten Merkmalsgruppen für jede Versuchsperson extrahiert. Darauffolgend wird jeweils der Abstand zum gelernten Modell berechnet. Diese Distanzen werden gemittelt, um eine allgemeine Abschätzung über alle Probanden zu erhalten. Je geringer die Distanz, desto mehr entspricht das Verhalten in diesem Prozessschritt dem zuvor geschätzten Modell.

Zur Visualisierung und qualitativen Analyse werden die Distanzen im weiteren Verlauf durch eine Farbverteilung repräsentiert. Für eine gesamte Prozessanalyse werden die einzelnen Distanzen pro Prozessschritt in zeitlicher Reihenfolge als Heatmaps visualisiert. Je dunkler der einzelne Schritt in der jeweiligen Heatmap dargestellt ist, desto größer ist die Distanz zum entsprechenden gelernten Modell. Mit dieser Art der Darstellung ist auch eine qualitative Analyse aufeinanderfolgender Schritte möglich und es können Schwierigkeiten im Prozessablauf aufgedeckt werden.

7.2 Auswertung

In diesem Abschnitt stelle ich die Ergebnisse des entwickelten unüberwachten Verfahrens vor. Dabei werde ich besonders unterschiedliche und multimodale Kombinationen der

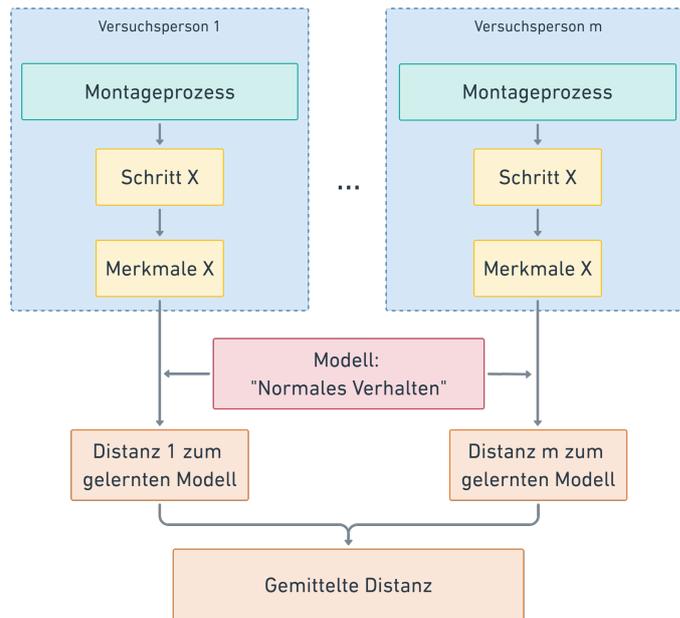


Abbildung 7.2.: Distanzberechnung mit dem gelernten Modell:

Bei der Betrachtung eines Prozessschrittes werden die Merkmale dieses Schrittes von jeder Versuchsperson extrahiert und die jeweiligen Distanzen zum gelernten Modell des „normalen Verhaltens“ berechnet. Abschließend werden diese Distanzen gemittelt.

Merkmalsgruppen analysieren, die die Basis der gelernten Gauß-Modelle des normalen Verhaltens sind.

7.2.1 Verwendung aller Merkmalsgruppen

In diesem Fall werden alle definierten Merkmalsgruppen und deren enthaltene Merkmale verwendet und durch ein einzelnes multidimensionalen Modell approximiert.

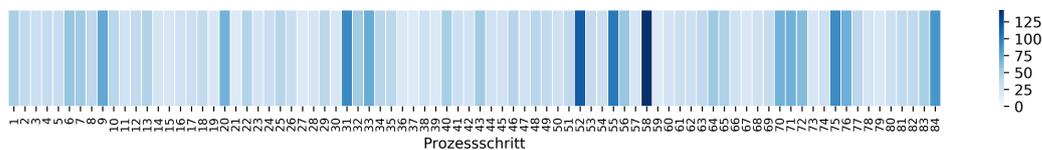


Abbildung 7.3.: Distanzen-Heatmap bei Verwendung aller Merkmalsgruppen:

Es ist das Mittel der Mahalanobis-Distanzen je Prozessschritt zum gelernten Modell „normalen Verhaltens“ (siehe Abbildung 7.2) zu sehen. Es werden alle Merkmalsgruppen verwendet. Je dunkler der Prozessschritt dargestellt ist, desto weiter ist dieser vom gelernten Modell entfernt.

Die entstehende Heatmap auf Basis der Distanzen der einzelnen Prozessschritte zum gelernten Modell ist in Abbildung 7.3 zu sehen. Im Folgenden werde ich auf auffällige Prozessschritte eingehen und diese qualitativ analysieren:



Abbildung 7.4.: 31. Schritt des Montageprozesses:

Es ist die durch das System gegebene Beschreibung des zu erfüllenden 31. Schrittes des Montageprozesses zu sehen.

Schritt 31

In diesem Prozessschritt (siehe Abbildung 7.4) erfolgte ein Wechsel des Arbeitsbereiches. Er unterscheidet sich deutlich zu einem normalen Montageschritt, da der Proband den Schritt deutlich schneller durch das Bewegen des Montageschlittens in den nächsten Arbeitsbereich beendet und kaum auf den Bildschirm schaut. Eine Erkennung dieses Schrittes durch das vorgestellte Verfahren ist aus Sicht der Methodik sinnvoll, da dieser deutlich von den eigentlichen Montageaufgaben abweicht. Dieser Schritt bietet jedoch kein Potential für eine Optimierung des Prozesses und müsste in zukünftigen Arbeiten gesondert behandelt bzw. maskiert werden.

Schritt 52

Dieser Prozessschritt (siehe Abbildung 7.5) beschreibt das Aufnehmen einer Mutter mit dem Druckluftschrauber. Diese Handlung wurde schon in Schritt 47 und 49 identisch durchgeführt. Beim Vergleich dieser drei Schritte wird deutlich, dass Schritt 52 weiter vom Modell entfernt ist als Schritt 49. Außerdem ist Schritt 49 weiter entfernt als Schritt 47. Daraus lässt sich schließen, dass Schritt 52 in dieser Form nicht nötig ist und angepasst werden sollte. Eine Erkennung dieses Schrittes ist also sinnvoll für eine Prozessoptimierung.

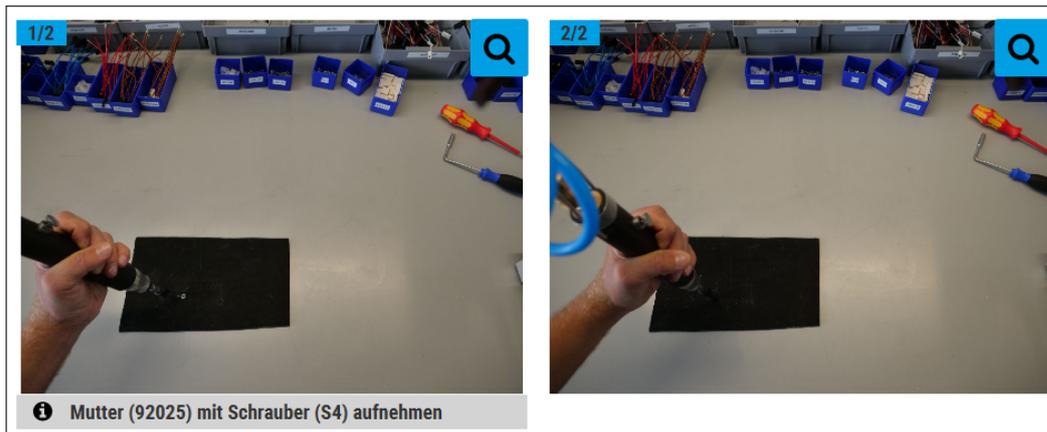


Abbildung 7.5.: 52. Schritt des Montageprozesses:

Es ist die durch das System gegebene Beschreibung des zu erfüllenden 52. Schrittes des Montageprozesses zu sehen.

Schritt 55

In diesem Schritt (siehe Abbildung 7.6) kommt es erneut zu einer starken Abweichung zum normalen Verhalten. Wie schon bei der Beschreibung von Schritt 52 genannt, kommt es zwischen den Schritten 47 und 56 zu mehrfachen Wiederholungen. Der Klemmkasten wird durch vier Fixierungen am Motor befestigt und jede Fixierung beinhaltet dieselbe Beschreibung und dieselben Abläufe. Es lässt sich nun vermuten, dass auch Schritt 55 ein Prozessschritt ist, der angepasst werden sollte, um die Wiederholungen in der Beschreibung zu vermeiden. Diese Schritte werden oftmals schnell bestätigt, da der Versuchsperson klar zu sein scheint, was zu tun ist.

Schritt 58

Prozessschritt 58 (siehe Abbildung 5.5, Seite 68) ist am weitesten vom geschätzten Modell entfernt. Dieser Schritt ist einer der manuell eingeführten schwerwiegenden Prozessstörungen (siehe Abschnitt 5.1.3, Seite 63). Die Versuchsperson realisiert in diesem Schritt den falschen Einbau eines Teils und die Beschreibung des Schrittes durch das Assistenzsystem passt nicht mehr zu der Situation in der realen Welt. Diese schwerwiegende Prozessstörung wurde somit erfolgreich durch das Verfahren erkannt.

Schritte 70–72 und Schritt 75–77

Diese beiden Sequenzen von Prozessschritten sind sich sehr ähnlich. In der Heatmap (siehe Abbildung 7.3) sind in diesen Zeiträumen deutliche Abweichungen zum gelernten Modell zu erkennen. In beiden Fällen müssen die Versuchspersonen über



Abbildung 7.6.: 55. Schritt des Montageprozesses:

Es ist die durch das System gegebene Beschreibung des zu erfüllenden 55. Schrittes des Montageprozesses zu sehen.

mehrere Prozessschritte hinweg Litzen durch Verdrillen miteinander verbinden, um diese im jeweils letzten Schritt mithilfe einer Rohrzange und einem Endverbinder zu fixieren. Auch in diesem Fall kann durch eine qualitative Analyse der Schritte gezeigt werden, dass eine Erkennung als Abweichung vom normalen Verhalten sinnvoll ist. Die in diesen Schritten zu erfüllenden Aufgaben sind im Gegensatz zu den restlichen Prozessschritten deutlich körperlicher. Das Verbinden der Litzen mithilfe des Endverbinders und der Rohrzange ist nur unter Anwendung von viel Kraft möglich. Der Endverbinder muss durch die Versuchsperson verbogen werden, damit dieser die Kabel zusammenhält. Bringt die Versuchsperson zu wenig Kraft auf, fällt der Endverbinder wieder ab. Diese Verbiegung konnten die Teilnehmenden meist nur nach mehreren Versuchen richtig durchführen. Somit ist der Schritt anspruchsvoller als die anderen Aufgaben. Im beschriebenen praktischen Szenario (siehe Kapitel 3, Seite 25) wird dieser Schritt mit einem druckluftbetriebenen Presswerkzeug durchgeführt, welches für die Datenerhebung jedoch nicht verwendet wurde (siehe Abschnitt 5.1.2, Seite 63). Auch in diesen Fällen ist somit der große Abstand zum normalen Verhaltensmodell gerechtfertigt.

Schritt 84

Auch das Verhalten der Probanden in diesem bereits zuvor diskutierten Schritt der Sichtprüfung (siehe Abschnitt 6.2.4, Seite 96) ist weit vom normalen Verhalten entfernt. Eine Erkennung dieser modellierten schwerwiegenden Prozessstörung ist, wie schon beschrieben, sinnvoll.

7.2.2 Verwendung der Gesichtsausdrücke

Im Gegensatz zu der multimodalen Analyse aller Merkmalsgruppen werden in diesem Abschnitt einzig die Merkmalsgruppen betrachtet, die Gesichtsausdrücke enthalten. Da es wie zuvor gezeigt keinen signifikanten Unterschied zwischen der Betrachtung der Intensität und des Auftretens von AU gibt, wird die Merkmalsgruppe AU-R und somit die Intensität verwendet.

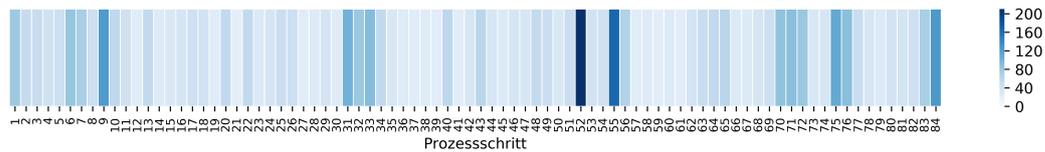


Abbildung 7.7.: Distanzen-Heatmap bei Verwendung der Action Units:

Es ist das Mittel der Mahalanobis-Distanzen je Prozessschritt zum gelernten Modell „normalen Verhaltens“ (siehe Abbildung 7.2) zu sehen. Es werden ausschließlich Intensitäten von Gesichtsausdrücken verwendet. Je dunkler der Prozessschritt dargestellt ist, desto weiter ist dieser vom gelernten Modell entfernt.

Zunächst werden alle Merkmale dieser Gruppe (somit alle AUs) zu einem multidimensionalen Modell zusammengefasst und nach dem beschriebenen Ablauf verarbeitet. Die Heatmap der multidimensionalen Gauß-Verteilung ist in Abbildung 7.7 dargestellt.

Erneut werde ich auf auffällige Prozessschritte qualitativ eingehen. Es kann davon ausgegangen werden, dass in dieser Analyse auch ähnliche Prozessschritte erkennbar sind wie bei dem zuvor beschriebenen Verfahren unter Berücksichtigung aller Merkmale, da die AUs auch in dem zuvor gelernten Modell enthalten sind.

Schritt 9

Dieser Schritt (siehe Abbildung 7.8) ist einer der am weitesten entfernten Schritte vom gelernten Modell. Er ist in der unüberwachten Analyse zuvor nicht aufgefallen. Dabei handelt es sich um die Anweisung, dass eine Unterlegscheibe auf eine Schraube gesteckt werden soll. Diese Aufgabe hat keine hohe Komplexität. Dennoch zeigen die Versuchspersonen eine Mimik, die nicht gut durch das gelernte Modell abgebildet werden kann. Eine mögliche Erklärung für dieses Verhalten ist, dass die gezeigten Bilder in der Anweisung sich nur sehr gering vom vorherigen Schritt unterscheiden, in dem die Schraube in das Loch gesteckt wurde. Die Schraube, der Klemmkasten und die zu platzierende Unterlegscheibe sind grau und können nur schlecht auf den Bildern erkannt werden. Außerdem gibt das System die Warnung,

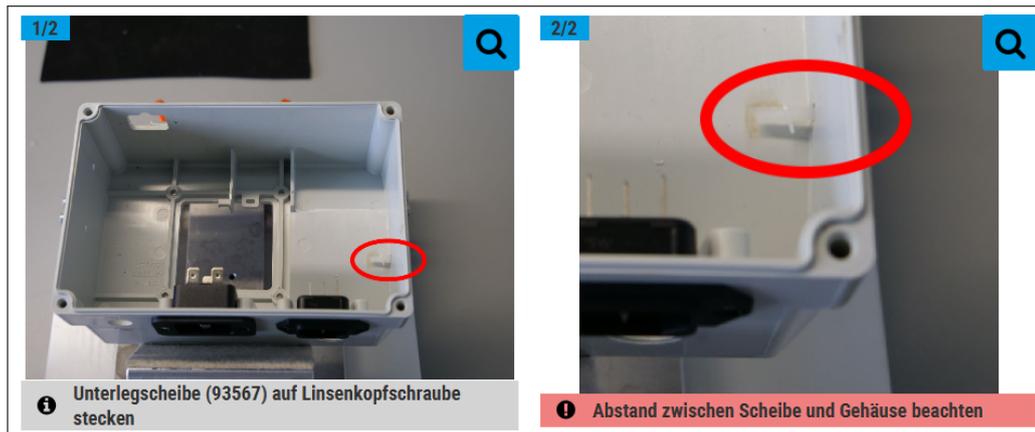


Abbildung 7.8.: 9. Schritt des Montageprozesses:

Es ist die durch das System gegebene Beschreibung des zu erfüllenden 9. Schrittes des Montageprozesses.

dass man den Abstand zwischen dem Gehäuse und der Scheibe beachten soll. Dieser Hinweis kann zu Verwirrung führen, da der Abstand nicht näher definiert wird und er auch auf der Abbildung nicht deutlich wird.

Schritt 52/55

Besonders Schritt 52 (siehe Abbildung 7.5, Seite 110) und Schritt 55 (siehe Abbildung 7.6, Seite 111) sind, wie auch in der Analyse zuvor, weit vom gelernten Modell entfernt. Wie oben beschrieben sind diese Schritte ein Teil sich wiederholender gleicher Anweisungen. Dieses scheint sich auch durch gezeigte AUs bei den Versuchspersonen widerzuspiegeln.

Schritt 58

Dieser Schritt hat bei der reinen Analyse der Mimik keinen großen Abstand zu dem gelernten Modell (im Gegensatz zu der multimodalen Analyse zuvor). Daraus lässt sich schließen, dass in diesem Fall andere Merkmalsgruppen relevant sind und die Mimik der Versuchspersonen normales Verhalten wiedergespiegelt hat. In diesem Falle scheint besonders die Dauer des Schrittes relevant zu sein (siehe unten).

Schritte 70–72, Schritte 75–77, Schritt 84

Es lassen sich erneut sowohl Schritt 84 (siehe Abbildung 5.6, Seite 68) als auch die Sequenzen Schritte 70–72 und Schritte 75–77 erkennen. Diese habe ich bereits zuvor beschrieben. Die Unterschiede auch bei einer reinen Betrachtung der Gesichtsausdrücke lassen darauf schließen, dass in diesen Schritten deutlich andere Gesichtsausdrücke gezeigt werden.

7.2.3 Verwendung der System-Events

Auch die Merkmalsgruppe der System-Events lässt sich durch ein multivariates Gauß-Modell approximieren. Die resultierende Heatmap der Distanzen ist in Abbildung 7.9 dargestellt. Hierbei fallen die folgenden Schritte besonders auf:

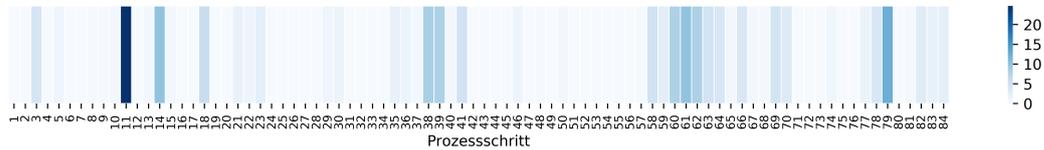


Abbildung 7.9.: Distanzen-Heatmap bei Verwendung der System-Events:

Es ist das Mittel der Mahalanobis-Distanzen je Prozessschritt zum gelernten Modell „normalen Verhaltens“ (siehe Abbildung 7.2) zu sehen. Es werden ausschließlich System-Events verwendet. Je dunkler der Prozessschritt dargestellt ist, desto weiter ist dieser vom gelernten Modell entfernt.

Schritt 11

Dieser Schritt beinhaltet eine der manuell eingeführten leichten Prozessstörungen (siehe Abschnitt 5.1.3, Seite 63). Diese leichte Störung scheint den Ablauf so zu beeinflussen, dass das System anders bedient wird als in den übrigen Schritten. Durch das Entfernen der Information aus dem Text könnten die Probanden die Bilder nun genauer betrachten, um die fehlende Information zu ermitteln.

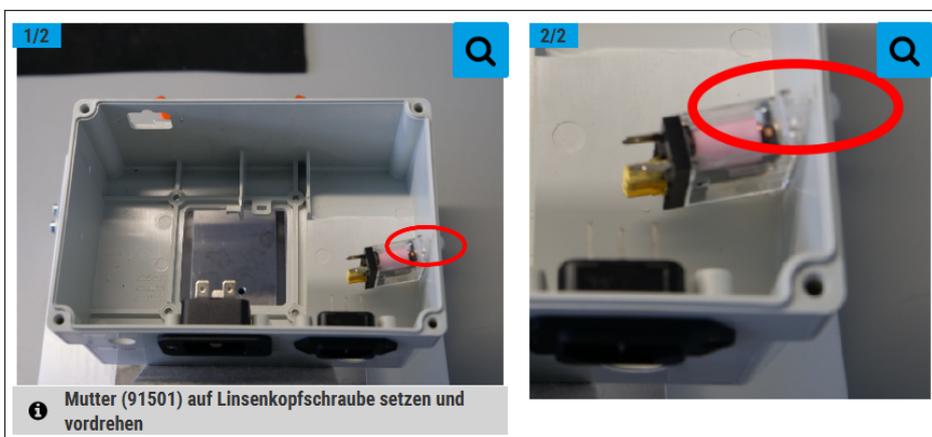


Abbildung 7.10.: 14. Schritt des Montageprozesses:

Es ist die durch das System gegebene Beschreibung des zu erfüllenden 14. Schrittes des Montageprozesses.

Schritt 14

In diesem Schritt muss eine Mutter befestigt werden (siehe Abbildung 7.10). Die

Bilder sind jedoch unscharf und auch die Mutter ist auf den Bildern kaum erkennbar. Die Probanden versuchen auch in diesem Fall durch Systeminteraktion mehr Informationen zu erhalten. Eine Erkennung dieses Schrittes ist demzufolge sinnvoll und die Bilder sollten bei der zukünftigen Verwendung des Systems erneuert werden.

Schritte 38–39

Diese Schritte sind Teil der Schutzleiterfixierung (bestehend aus fünf Einzelschritten (Schritte 36–40)). Dabei wird in mehreren Schritten eine anspruchsvolle Aufgabe erledigt und nacheinander Teile freihändig zu einem Schutzleiter zusammengefügt. Dieser wird anschließend am Motor befestigt. Die einzelnen Schritte sind in einer Art und Weise beschrieben, dass man die zusammengefügte Teile immer in den Händen hält. Dabei ist aber die Touch-Interaktion mit dem System schwierig. Es kann dazu kommen, dass die zusammengesetzten Teile vor der finalen Fixierung am Motor wieder auseinanderfallen und erneut mit dem Zusammenbau begonnen werden muss. Somit greift der Proband häufig auf die Dokumentation zurück, um sich die vorherigen Schritte erneut anzuschauen.

Schritte 58–64

In diesem Bereich häufen sich die manuell hinzugefügten Prozessstörungen. Dieses kann durch die Abweichung vom gelernten Modell erkannt werden. Es scheint also zu einer abweichenden direkten Systeminteraktion zu kommen, wenn Prozessstörungen auftreten.

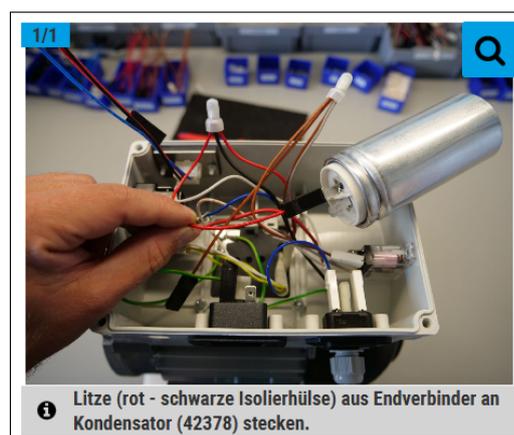


Abbildung 7.11.: 79. Schritt des Montageprozesses:

Es ist die durch das System gegebene Beschreibung des zu erfüllenden 79. Schrittes des Montageprozesses.

Schritt 79

In diesem Schritt muss eine Litze an einem Kondensator befestigt werden. Die dazugehörige Beschreibung ist in Abbildung 7.11 zu sehen. Es kann hierbei zu Verwirrung kommen, da nicht klar ist, welches der roten Kabel aus dem Endverbinder verwendet werden soll. Aus diesem Grund wird mit dem System interagiert und das Bild wird genauer angeschaut.

7.2.4 Verwendung der Aufmerksamkeit

Bei der alleinigen Berücksichtigung der Aufmerksamkeit als Basis eines multidimensionalen Gauß-Modells können weitere Erkenntnisse über den Prozess gewonnen werden.

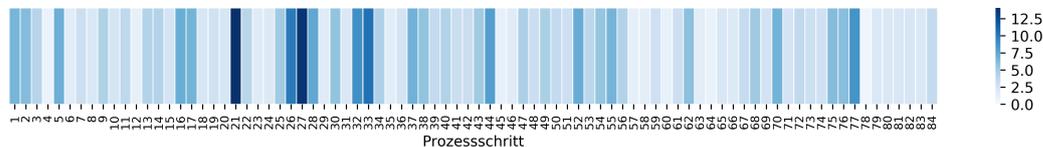


Abbildung 7.12.: Distanzen-Heatmap bei Verwendung der Aufmerksamkeit:

Es ist das Mittel der Mahalanobis-Distanzen je Prozessschritt zum gelernten Modell „normalen Verhaltens“ (siehe Abbildung 7.2) zu sehen. Es wird ausschließlich Aufmerksamkeit verwendet. Je dunkler der Prozessschritt dargestellt ist, desto weiter ist dieser vom gelernten Modell entfernt.

Die Heatmap ist in Abbildung 7.12 zu sehen. Die Distanzen zum gelernten Modell sind kleiner als bei den zuvor betrachteten Merkmalsgruppen. Die Schritte unterscheiden sich also in Hinblick auf diese Merkmalsgruppe nicht so stark vom gelernten Modell. Auch in diesem Abschnitt werde ich auffallenden Prozessschritte qualitativ beschreiben:

Schritte 1–3

Es ist zu beobachten, dass zu Beginn des Versuches ein Unterschied in der Aufmerksamkeit zu sehen ist. Dieses lässt sich durch einen Neuheitseffekt erklären. Die Versuchspersonen müssen sich zunächst an das System gewöhnen und die relevanten Bereiche der modellierten Oberfläche verinnerlichen. Wie in Abbildung 7.12 zu sehen ist, nimmt dieser Effekt ab Schritt 3 jedoch ab.

Schritt 21

Dieser Schritt beinhaltet die Fixierung einer Schraube in identischer Art und Weise, wie es auch in Schritt 20 der Fall ist. Aus diesem Grund lässt sich vermuten,

dass der Aufmerksamkeitsfokus besonders auf dem Werkstück und nicht auf den Bildschirmen liegt, da bereits klar ist, wie die Aufgabe zu lösen ist.

Schritte 25–28

Auf diese sich wiederholenden Prozessschritte der Befestigung des Anlaufrelais bin ich bereits zuvor eingegangen. Auch bei einer alleinigen Betrachtung der Aufmerksamkeit kann in diesem Bereich eine Auffälligkeit erkannt werden. Gerade die hohe Distanz in Schritt 27 ist auffällig. Die Mehrzahl der Versuchspersonen wissen bereits, was sie zu tun haben und ihre Aufmerksamkeit liegt nicht mehr auf dem System sondern nur noch auf dem Werkstück. Entweder haben sie das Relais bereits fixiert und bestätigen die Schritte nachträglich oder sie müssen keinen Abgleich mehr zwischen virtueller und realer Welt durchführen, da sie bereits wissen, was zu tun ist, um die Aufgabe zu lösen.

Schritte 32–34

In diesen Schritten wird der Motorschlitten gedreht. Ich vermute, dass die Versuchspersonen zunächst verstehen müssen, wie der Schlitten gedreht werden soll und wie dessen Position gerade in der realen Welt ist. In diesem Schritt wird die Drehfunktion erstmalig referenziert. Da diese Funktion zuvor keine Rolle gespielt hat, muss der Proband zunächst verstehen, wie die Anweisung überhaupt auszuführen ist und zeigt somit ein anderes Aufmerksamkeitsverhalten.

Schritte 75–77

Wie bereits zuvor beschrieben, muss in diesen Schritten eine komplexere Aufgabensequenz (Verbinden von Litzen durch Verdrillen) durchgeführt werden. Es wird deutlich, dass dabei auch der Aufmerksamkeitsfokus ein nicht normales Verhalten zeigt. Dieses wird vor allem daran liegen, dass die komplexe Aufgabe schwieriger zu verstehen ist und somit die Aufmerksamkeitsmerkmale anders sind als bei den restlichen zumeist leicht verständlichen Schritten.

7.2.5 Verwendung der Dauer

Bei alleiniger Berücksichtigung der Dauer der einzelnen Schritte und der Modellierung der Gauß-Verteilung auf Grundlage dieses Merkmals, kommt es bei der Betrachtung der Distanzen zu der in Abbildung 7.13 dargestellten Heatmap. Dabei fällt besonders ein einzelner Schritt auf:

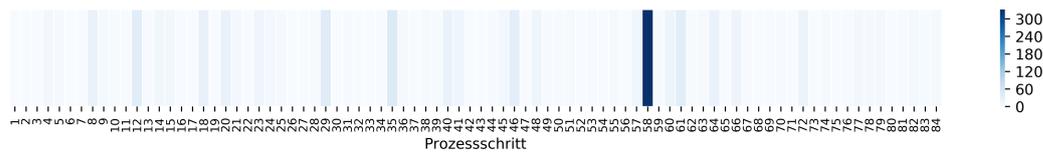


Abbildung 7.13.: Distanzen-Heatmap bei Verwendung der Dauer:

Es ist das Mittel der Mahalanobis-Distanzen je Prozessschritt zum gelernten Modell „normalen Verhaltens“ (siehe Abbildung 7.2) zu sehen. Es wird ausschließlich Dauer verwendet. Je dunkler der Prozessschritt dargestellt ist, desto weiter ist dieser vom gelernten Modell entfernt.

Schritt 58

Besonders dieser Prozessschritt (siehe Abbildung 5.5, Seite 68) weist bei einer reinen Betrachtung der Prozessschrittdauer eine große Distanz zum gelernten Modell auf. Dieses ist, wie bereits zuvor beschrieben, durch den unbewusst falschen Einbau eines Teiles zu erklären, welches in diesem Schritt referenziert wird. Dieser Prozessschritt dauert daher deutlich länger als die restlichen Schritte, da die Versuchspersonen zumeist den Fehler beheben und das Teil ausbauen und erneut einbauen.

Schritt 35

Schritt 35 hat den zweitgrößten Abstand zum Modell (siehe Abbildung 5.3, Seite 66). Dabei ist zu beachten, dass Schritt 58 mehr als 7 mal weiter vom Modell entfernt ist. Dieser Prozessschritt beinhaltet eine weitere manuell hinzugefügten schwerwiegende Prozessstörung. Es sind bereits Teile montiert, die in diesem Schritt vom Probanden erst hätten montiert werden sollen. Die Dauer dieses Schrittes ist dennoch deutlich anders als im gelernten Modell und die Prozessstörung kann dadurch erkannt werden.

7.3 Zusammenfassung und Diskussion

In diesem Kapitel habe ich ein Verfahren vorgestellt, dass in der Lage ist unüberwachte Prozesse zu analysieren und Verbesserungspotentiale in einzelnen Prozessschritten anhand menschlichen Verhaltensweisen aufzudecken. Nach der Beschreibung des entwickelten Systems habe ich eine multimodale Analyse auf Grundlage der Verwendung aller Merkmalsgruppen qualitativ analysiert. Darauf folgend habe ich jeweils die einzelnen Merkmalsgruppen (Gesichtsausdrücke, System-Events, Aufmerksamkeit, Dauer) betrachtet und Unterschiede und Gemeinsamkeiten aufgedeckt und beschrieben.

Der in diesem Kapitel vorgestellte unüberwachte Ansatz ist in der Lage einem Prozessentwickler als Werkzeug zu dienen, um in dem entwickelten Prozess durch die Analyse des menschlichen Verhaltens bei der Systembedienung Potentiale für Verbesserungen aufzudecken.

Die reine Betrachtung der Dauer der einzelnen Prozessschritte für eine Bewertung ist nicht die beste Lösung. Es können zwar Ergebnisse erzielt werden, bei einer Berücksichtigung zusätzlicher Merkmale wie des Gesichtsausdruckes, können aber weitere Potentiale aufgedeckt werden.

Im vorgestellten unüberwachten Verfahren wird das Modell durch den gesamten Datensatz trainiert. Die von mir gemachten Analysen basieren auf der Distanz einzelner Schritte zu dem gelernten Modell. Die diskutierten Prozessschritte in der Auswertung sind demzufolge bereits im Training enthalten. Diese Unterteilung anhand einzelner Prozessschritte sollte nicht als klassisches Klassifikationsverfahren angesehen werden. Es ist vielmehr eine Analyse der Gesamtheit der vorhandenen Daten in Hinblick auf einen einzelnen Schritt. Bildlich gesprochen betrachte ich, wie gut die Daten eines Prozessschrittes vom gelernten Modell abgebildet werden. Es gibt so gesehen keine klassische Trennung zwischen Test- und Trainingsdaten, sondern einen Datensatz, der zur genaueren Analyse unterschiedlich gefiltert wird.

Der unüberwachte Ansatz ist nicht in der Lage alle künstlich hinzugefügten schwerwiegenden Prozessstörungen zu erkennen. Auch die leichten Prozessstörungen werden zumeist nicht durch das Modell erkannt. Dieses kann bei den leichten Prozessstörungen – wie bereits im Kapitel zuvor beschrieben – daran liegen, dass die Versuchspersonen den Schritt trotz der Änderungen mit einer normalen Verhaltensweise lösen können. Sie benötigen keine zusätzlichen Informationen, um die Aufgabe zu erledigen und somit gibt es auch keine Störung im Prozess. Ich bin davon ausgegangen, dass die modellierten schwerwiegenden Prozessstörungen so gravierend sind, dass in jedem Falle mit einer Reaktion im Verhalten der Teilnehmenden zu rechnen ist. Nun ist diese starke Reaktion jedoch bei einigen dieser Schritte ausgeblieben oder die gewählten Merkmalsgruppen können die Reaktion nicht erfassen.

Dennoch ist das beschriebene Verfahren in der Lage, Auffälligkeiten im Prozess anhand menschlichen Verhaltens zu erkennen. Dieses konnte ich nicht zuletzt durch die qualitative Analyse zeigen. Somit kann ich Hypothese 7.1 (Durch unüberwachte Verfahren können Prozessstörungen detektiert werden.) bestätigen.

Auch Hypothese 7.2 (Die Betrachtung von multimodaler Merkmalsgruppenkombinationen hat einen Mehrwert gegenüber der reinen zeitlichen Analyse.) ist erfüllt, da klar wird, dass eine multimodale Modellierung weitere und detailliertere Erkenntnisse liefert. Besonders beim Vergleich zu einer reinen zeitlichen Analyse fällt dieses auf. Dennoch hat auch die Analyse einzelner Merkmalsdimensionen ihre Berechtigung. So können durch die reine Betrachtung der einzelnen Merkmale zusätzliche Besonderheiten erkannt werden (wie zum Beispiel Schritt 58 in der zeitlichen Analyse). Ich schlage somit als optimale Lösung eine Kombination aus multimodalen und unimodalen Merkmalsanalysen vor.

Die Heatmaps kann ein Prozessmodellierer nutzen, um sich anhand der objektiven Evaluationsdaten eine Meinung über den Prozess zu bilden. Diese Art der Darstellung eignet sich besonders, um einen Überblick über den gesamten Prozess und etwaige Potentiale zu erhalten. Somit ist die Grundlage geschaffen, um Hypothese 7.3 (Es lässt sich ein Werkzeug entwickeln, das Prozessentwickler bei der Optimierung von Prozessabläufen unterstützt.) zu bestätigen. Auf Basis der unüberwachten Analyse kann ein Werkzeug entwickelt werden, das eine objektive Möglichkeit der Prozessoptimierung liefert.

Methodentransfer

“ *Things don't have to change the world to be important.*

– Steve Jobs

Hinweis Teile dieses Kapitels sind bereits von Harteis u. a. [Har+18] beschrieben und veröffentlicht worden.

In den vorherigen Kapiteln konnte ich zeigen, dass sowohl das entwickelte überwachte Analysesystem als auch der unüberwachte Ansatz im beschriebenen Szenario anwendbar sind und gute Ergebnisse liefern.

Um die Generalisierbarkeit der beschriebenen Ansätze zu zeigen, habe ich eine weitere Studie in Zusammenarbeit mit der Arbeitsgruppe „Erziehungswissenschaft mit dem Schwerpunkt Bildungsmanagement und Bildungsforschung in der Weiterbildung“ der Universität Paderborn konzipiert und umgesetzt. Die in dieser Arbeit beschriebenen überwachten (siehe Kapitel 6, Seite 77) und unüberwachten (siehe Kapitel 7, Seite 105) Verfahren werden dabei in eine vollständig andere Domäne übertragen und die folgenden Hypothesen betrachtet:

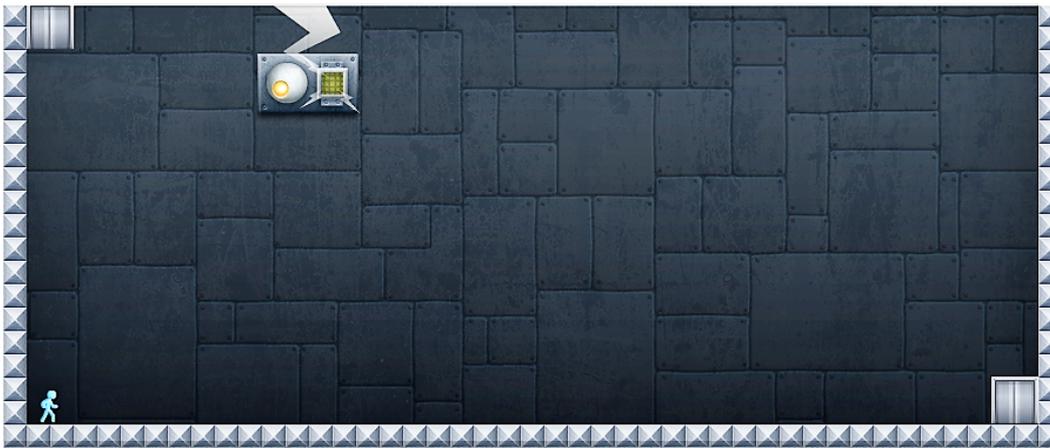
Hypothese 8.1 Menschen zeigen bereits vor dem Auftreten eines Fehlers ein Kopfverhalten, das den Fehler vorhersagbar macht.

Hypothese 8.2 Anhand des Kopfverhaltens kann die steigende Komplexität und Schwierigkeit im Verlauf der Studie erkannt werden.

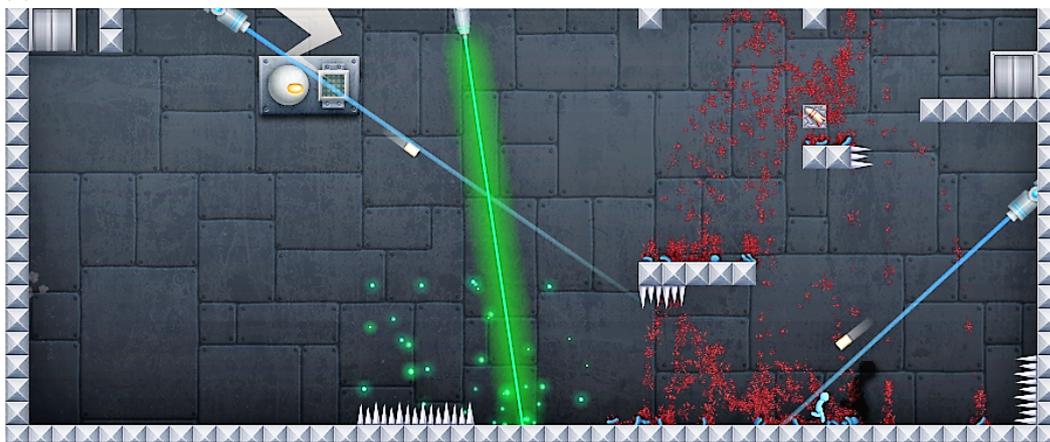
Für eine genauere Beschreibung und weitere Ergebnisse der Studie wie beispielsweise die Verbindung zwischen Lernerfolg und Mimik referenziere ich auf die Veröffentlichung von Harteis u. a. [Har+18]. In diesem Kapitel werde ich Aspekte darstellen, die die Thematik dieser Arbeit betreffen.

8.1 Studienbeschreibung

Um die Hypothesen betrachten zu können, wurde ein kontrollierbares Szenario gewählt. Die Auswahl ist dabei auf das Spiel „Give Up 2“ von Tasselfoot und Armor Games¹ gefallen. Es ist ein Jump ‘n’ Run Spiel, in dem der Spieler eine Figur steuert, die immer mehr und schwierigere Hindernisse überwinden muss, um die Ausgangstür des Levels zu erreichen und dieses somit erfolgreich abzuschließen.



(a) Level 1: Keine Hindernisse vorhanden



(b) Level 8: Eine Vielzahl von tödlichen (z. B. Kanonen) und nicht-tödlichen Hindernissen (z. B. Plattformen) vorhanden

Abbildung 8.1.: Vergleich zwischen Level 1 und Level 8 des Spiels „Give Up 2“:

Es ist der Vergleich zweier Level des Spiels „Give Up 2“ zu sehen. Für einen erfolgreichen Abschluss sind in Level 8 deutlich mehr Hindernisse zu überwinden als in Level 1.

¹<https://armorgames.com/play/17649/give-up-2>

Zwei Level sind beispielhaft in Abbildung 8.1 zu sehen. Dabei zeigt Abbildung 8.1a das erste Level. Es beinhaltet keine Hindernisse. Dem entgegengesetzt beinhaltet Level 8 (siehe Abbildung 8.1b) deutlich mehr Hindernisse und ist deutlich schwerer zu lösen.

Die Versuchsperson kann zwei unterschiedliche Fehler in dem Spiel begehen. Zum einen kann die Spielfigur durch Waffen und Hindernisse wie Stacheln getötet werden. Nach einem Tod muss die Versuchsperson das jeweilige Level erneut von Beginn starten. Der Tod wird im Spiel durch ein *blutiges Zerplatzen* der Spielfigur dargestellt. Die Vielzahl der vorherigen Tode ist auch in Abbildung 8.1b durch das viele dargestellte Blut zu erkennen. In Abschnitt 8.3 (siehe Seite 126) werde ich erneut auf das brutale Sterben der Spielfigur eingehen. Zum anderen kann die Versuchsperson daran scheitern Plattformen zu erreichen. Dabei stirbt die Figur nicht, die Versuchsperson muss jedoch erneut versuchen das Hindernis zu überwinden.

Da jede Versuchsperson die gleiche Abfolge an Level spielt, sind die Bedingungen bei jedem Studiendurchlauf gleich. Während des Spiels wird sowohl der Bildschirm als auch das Gesicht der Versuchspersonen aufgenommen. Neben der Analyse des Kopfausdruckes ist durch ein zusätzlich verwendetes Eye-Tracking Systems (SMI RED250) eine Analyse des Pupillendurchmessers und somit der mentalen Belastung [SKH17] möglich. Die Umgebungsbeleuchtung wurde durch künstliches Licht gleichgehalten. Die Versuchspersonen haben das Spiel für eine Dauer von fünf Minuten gespielt. In der folgenden Analyse werden ausschließlich Level betrachtet, die innerhalb dieser fünf Minuten erfolgreich durch die jeweilige Versuchsperson abgeschlossen wurden.

Die Figur lässt sich durch drei Tasten auf der Tastatur steuern. Sie kann nach links und rechts gelenkt werden. Darüber hinaus ist ein Sprung sowie ein Doppelsprung (zweifache Betätigung der Sprungtaste) möglich.

Insgesamt haben 38 Personen (23 weiblich, 15 männlich) an der Studie teilgenommen. Sie waren im Schnitt 29.79 Jahre alt (SD 5.92). Die jüngste Testperson hatte ein Alter von 20 Jahren, die älteste Testperson ein Alter von 45 Jahren.

Da in dieser Studie der Fokus auf eine Analyse des Gesichtsausdruckes gelegt wurde, wurden die Versuchspersonen angewiesen ihren Kopf bei der Studiendurchführung auf eine Kinnstütze abzulegen. Somit wurde verhindert, dass der Kopf bewegt und dadurch das Risiko von Fehlern in der Extraktion der Mimik erhöht wird.

Die Studie wurde durch die Ethik-Kommission der Universität Bielefeld genehmigt (siehe Anhang B.1.1, Seite 155). Den Teilnehmenden wurde die Studienbeschreibung (siehe Anhang B.1.2, Seite 156) ausgehändigt und sie unterzeichneten eine Einverständniserklärung (siehe Anhang B.1.3, Seite 157). Danach wurde ihnen das Spielprinzip in einem Demonstrationsvideo erklärt und es bestand die Möglichkeit der Klärung offener Fragen mit dem Versuchsleiter. Schließlich wurde die Studie gestartet und die Versuchspersonen spielten für 5 Minuten das beschriebene Spiel. Anschließend wurden sie gebeten, einen Persönlichkeitsfragebogen und einen Fragebogen zu ihrer persönlichen Einstellung gegenüber Computerspielen auszufüllen (siehe Anhang B.1.4, Seite 159).

Nach der Studiendurchführung wurden die aufgezeichneten Spieldaten manuell annotiert. Dabei wurden sowohl die beschriebenen Fehlerarten als auch Levelbeginn und Levelende zeitlich erfasst.

8.2 Anwendung der überwachten Verfahren

Das in Kapitel 6 (siehe Seite 77) von mir beschriebene Verfahren wird verwendet, um Hypothese 8.1 (Menschen zeigen bereits vor dem Auftreten eines Fehlers ein Kopfverhalten, das den Fehler vorhersagbar macht.) zu untersuchen. Neben den zuvor benutzten AUs werden in diesem Fall weitere Merkmale wie Valenz und Engagement verwendet. Ein zusätzlich verwendetes Merkmal ist der Pupillendurchmesser zur Schätzung der kognitiven Belastung. Außerdem wird die Blickrichtung nicht durch Aufmerksamkeitsbereiche approximiert, sondern sowohl der Blickwinkel als auch die Kopforientierung werden direkt verwendet. Harteis u. a. [Har+18] beschreiben die verwendeten Merkmale genauer.

Im Gegensatz zu der Merkmalsgenerierung zuvor, wird in diesem Fall nicht ausschließlich der Mittelwert im zu betrachtenden Zeitabschnitt genutzt, sondern darüber hinaus wird dieser Zeitabschnitt durch eine lineare Regression, einen Maximalwert, einen Minimalwert und der Standardabweichung approximiert. Somit ist der Merkmalsraum deutlich größer als im beschriebenen Montageszenario.

Dennoch kann dieselbe Verarbeitungskette angewendet und erneut ein binäres Klassifikationssystem modelliert werden. Dabei stellt sich vor allem die Frage, ob maschinelles Lernen in der Lage ist, Zeiträume vor einem Fehler von beliebigen anderen Zeiträumen zu unterscheiden. Aus diesem Grund werden Zeitfenster (Länge von 300 Millisekunden) direkt vor dem Auftreten eines Fehlers betrachtet. Diese Daten repräsentieren die positive

Klasse. Als negative Klasse werden alle Videos in Abschnitte mit einer Länge von 300 Millisekunden und einer Überlappung von 150 Millisekunden unterteilt. Dabei werden die 300 Millisekunden vor und nach dem gemachten Fehler nicht berücksichtigt. Die Datenmenge wird in Test- (25%) und Trainingsset (75%) unterteilt und das entwickelte Optimierungsverfahren (siehe Abschnitt 6.1.1, Seite 78) angewendet.

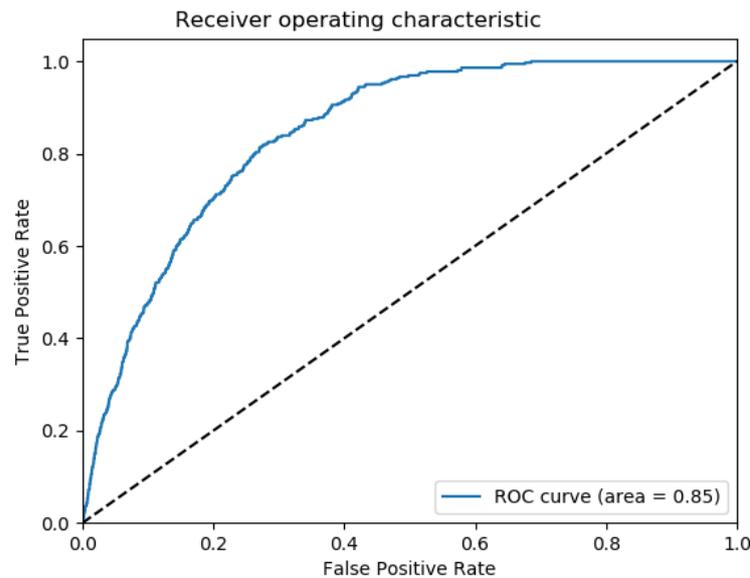


Abbildung 8.2.: ROC-Diagramm der besten Klassifikationskette:

Es ist das ROC-Diagramm der besten Klassifikationskette dargestellt. Diese Kette erzielt einen AUROC-Wert (area) von 0.85 und ist mit diesem gutem Leistungswert deutlich besser als eine Zufallsentscheidung (gestrichelte Linie).

Für die Optimierung wurden alle in Abschnitt 6.1.1 (siehe Seite 78) beschriebenen Klassifikatoren und Vorverarbeitungsschritte in eine Optimierung zusammengefasst. Die beste Leistung kann im Gegensatz zum Montage-Szenario eine boosting-basierte Klassifikationskette erreichen. In Abbildung 8.2 ist das ROC Diagramm der optimalen Klassifikationskette zu sehen. Mit einem AUROC-Wert von 0.85 kann die Klassifikation eine gute Leistung (Schulnote 2 (engl. B)) erzielen.

Somit ist das System in der Lage Bereiche vor gemachten Fehlern von den restlichen Bereichen zu unterscheiden. Damit kann ich Hypothese 8.1 bestätigen. Die Versuchspersonen zeigen im Zeitraum vor einem gemachten Fehler ein anderes Verhalten und der Fehler ist somit vorhersagbar. Dieses gute Ergebnis des überwachten Klassifikationssystems in

einem völlig anderen Anwendungsfall lässt darauf schließen, dass der Transfer in andere Bereiche möglich und der beschriebene Ansatz allgemein nutzbar ist.

8.3 Anwendung der unüberwachten Verfahren

Auch den beschriebenen unüberwachten Ansatz (siehe Kapitel 7, Seite 105) wende ich auf diese Studiendaten an. Da in diesem Szenario besonders die Gesichtsausdrücke im Mittelpunkt der Analyse stehen, werde ich mich bei dem Transfer des unüberwachten Verfahrens auf die Merkmalsgruppe AU-K (zeitlich normierte Auftrittserkennung der einzelnen AUs) beschränken. Im Vergleich dazu werde ich auch die Versuche pro Level mit dem unüberwachten Ansatz analysieren.

Die zu lösenden Aufgaben in dem betrachteten Computerspiel werden in jedem neuen Level schwieriger und komplizierter. Die Frage ist nun, ob auch durch eine direkte Analyse der Daten, diese Änderung der Levelkomplexität im Verhalten der Versuchspersonen erkannt werden kann und es besonders in den schweren Level zu einer anderen Verhaltensweise kommt. Im Gegensatz zum überwachten Ansatz im Abschnitt zuvor wird in diesem Fall jeweils das gesamte Level betrachtet. Ein Level ist damit analog zu einem Prozessschritt im zuvor beschriebenen Montageszenario (siehe Kapitel 3, Seite 25) zu sehen.

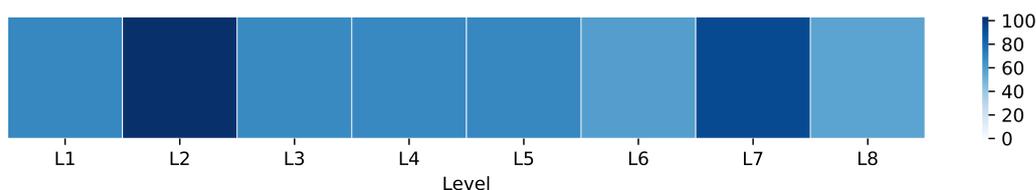


Abbildung 8.3.: Distanzen-Heatmap bei Verwendung der Action Units:

Es ist das Mittel der Mahalanobis-Distanzen je Prozessschritt zum gelernten Modell „normalen Verhaltens“ (siehe Abbildung 7.2) zu sehen. Es werden ausschließlich Auftritte von Action Units verwendet. Je dunkler der Prozessschritt dargestellt ist, desto weiter ist dieser vom gelernten Modell entfernt.

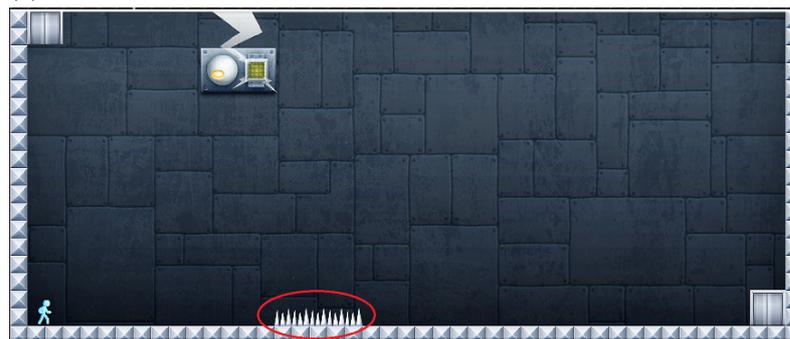
Es wird ein multidimensionales Gauß-Modell für das normale Verhalten in sämtlichen Leveln, wie in Kapitel 7 (siehe Seite 105) beschrieben, approximiert. Darauffolgend werden

wiederum die Mahalanobis-Distanzen pro Level berechnet und gemittelt, um ein Abstandsmaß zu dem gelernten Verhaltensmodell zu erlangen. Die Distanzen sind in Abbildung 8.3 erneut als Heatmap zu sehen.

Insgesamt haben die einzelnen Versuchspersonen maximal neun Level erfolgreich abgeschlossen. Nicht jede Versuchsperson hat das neunte Level erreicht, da die gesamte Studiendauer von fünf Minuten zuvor abgelaufen war. Im Folgenden werde ich qualitativ auf die Auffälligkeiten der Ergebnisse des unüberwachten Ansatzes eingehen:



(a) Level 1: Keine Hindernisse vorhanden



(b) Level 2: Erstmaliges Vorkommen eines tödlichen Hindernisses (siehe Stacheln in der roten Markierung)

Abbildung 8.4.: Vergleich zwischen Level 1 und Level 2 des Spiels „Give Up 2“:

Nach erfolgreichem Abschluss des ersten Levels kommt im zweiten Level erstmalig ein tödliches Hindernis vor.

Level 2

In Level 2 ist eine eindeutige Abweichung vom normalen mimikbasierten Verhalten feststellbar. Der Unterschied zwischen Level 1 und Level 2 ist in Abbildung 8.4 zu sehen. Im zweiten Level ist zum ersten Mal ein Hindernis vorhanden, dass die Spielfigur töten kann (siehe rot markierte Stacheln in Abbildung 8.4b). Dadurch kommt es in diesem Level für viele der Versuchspersonen zum ersten Tod und sie sehen das *blutige Zerplatzen* der Spielfigur zum ersten Mal. Ich vermute, dass beim

ersten Tod deutlich anderes Verhalten gezeigt wird als bei darauffolgenden Toden, da sich die Versuchspersonen zunehmend an die visuellen Reize gewöhnen.



(a) Level 6: Keine permanenten tödlichen Bedrohungen vorhanden



(b) Level 7: Erstmaliges Vorkommen von permanenten tödlichen Bedrohungen (siehe Kanonen in der roten Markierung)

Abbildung 8.5.: Vergleich zwischen Level 6 und Level 7 des Spiels „Give Up 2“:

Nach erfolgreichem Abschluss des sechsten Levels kommen im siebten Level erstmalig permanente tödliche Bedrohungen hinzu. Die Spielfigur wird getötet, wenn sich diese nicht bewegt.

Level 7

Auch in Level 7 wurde ein deutlich anderes Verhalten gezeigt als im gelernten Modell abgebildet ist. In diesem Level tritt erstmalig eine permanente Bedrohung auf. Level 6 und Level 7 sind in Abbildung 8.5 zu sehen. Als neue Schwierigkeit kommen in Level 7 zwei Kanonen hinzu, die auf die Spielfigur schießen (siehe rote Markierung in Abbildung 8.5b). Der blaue Strahl ist die Zielvorrichtung. Befindet sich die Spielfigur in Reichweite der Kanone, feuert diese in definierten Abständen in ihre Richtung. Die Versuchsperson muss permanent den Geschossen ausweichen und dennoch die Tür zum erfolgreichen Abschluss des Levels erreichen. Diese Kanonen sind im Vergleich zu den Anpassungen zuvor ein deutlicher Anstieg der

Schwierigkeit. Dieser Anstieg ist eindeutig in der unüberwachten Analyse der Mimik zu sehen.

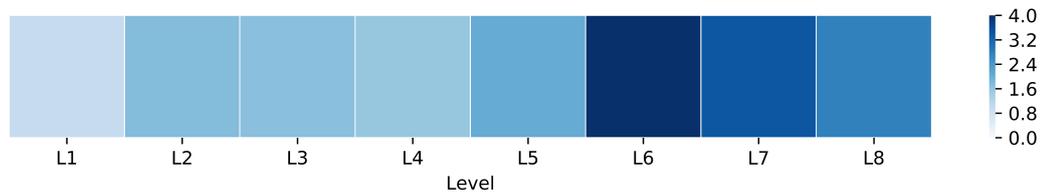


Abbildung 8.6.: Distanzen-Heatmap bei Verwendung der Levelversuche:

Es ist das Mittel der Mahalanobis-Distanzen je Prozessschritt zum gelernten Modell „normalen Verhaltens“ (siehe Abbildung 7.2) zu sehen. Es werden ausschließlich Levelversuche verwendet. Je dunkler der Prozessschritt dargestellt ist, desto weiter ist dieser vom gelernten Modell entfernt.

Im Vergleich zu einer multidimensionalen Modellierung auf Basis von AUs wurde der unüberwachte Ansatz auch auf die Anzahl der Versuche pro Level angewendet. Wiederum wurde ein normales Verhalten geschätzt, um dieses mit dem Verhalten in den einzelnen Level zu vergleichen. Abbildung 8.6 zeigt das Ergebnis und ich gehe auf die auffälligen Level im Folgenden ein:

Level 6

In Level 6 ist ein deutlicher Unterschied zum normalen Verhalten erkennbar. Dieses wird durch die zuvor beschriebene Modellierung auf Basis der AUs Analyse nicht so deutlich erkannt. Demzufolge wiederholen die Versuchspersonen das Level zwar häufiger als normal, sie zeigen aber keine gravierend abweichende Mimik.

Level 7–8

Wie auch schon bei der Verwendung der Mimik zuvor, ist die gestiegene Schwierigkeit bei einem Fortschreiten der Level deutlich zu sehen. Die hinzugekommene permanente Bedrohung in Level 7 (siehe Abbildung 8.5) ist auch in der Abweichung der Anzahl der Versuche zu erkennen.

Es fällt auf, dass Level 6 bei der unüberwachten Analyse auf Basis der Versuche weiter vom gelernten Modell entfernt ist als Level 7. Dieses widerspricht auf dem ersten Blick der zuvor getroffenen Annahme, dass Level 7 deutlich auffallen müsste, da erstmalig eine permanente Bedrohung auftritt. Dieser Widerspruch kann dadurch erklärt werden, dass viele Versuchspersonen Level 7 nicht erfolgreich abschließen können, da die maximale Studiendauer erreicht wurde. Somit sind nur noch wenige Beispiele der letzten erreichten

Level 7 und Level 8 im Datensatz vorhanden. Ein Level wird nicht berücksichtigt, wenn das Level vor dem Ende der fünf Minuten nicht erfolgreich beendet wurde. Somit wird eine potentiell hohe Anzahl an Versuchen in diesen schwierigen Level in der Auswertung nicht einbezogen. Dennoch ist bemerkenswert, dass in Level 7 deutlich andere mimikbasierte Verhaltensweisen gezeigt werden und somit durch die Analyse der Mimik zusätzliche Informationen aus der Analyse abgeleitet werden können. Level 9 wird nicht in der Auswertung betrachtet, da zu wenige der betrachteten Versuchspersonen dieses Level erfolgreich in der vorgegebenen Zeit beendet haben.

8.4 Zusammenfassung und Diskussion

In diesem Kapitel habe ich gezeigt, dass die entwickelten Verfahren der überwachten Klassifikation und der unüberwachten Analyse auch in anderen Bereichen und Szenarien anwendbar sind. Dabei konnte ich durch die unterschiedlichen zur Verfügung stehenden Merkmale die Konzepte übertragen. Allerdings musste ich die Verarbeitungsketten der Klassifikation und auch die unüberwachten Modelle neu trainieren. Ein Transfer der in der Montageumgebung gelernten Verarbeitungsketten und Modelle in die Spielumgebung aus diesem Kapitel wurde nicht durchgeführt. Ich konnte aber zeigen, dass in beiden dargestellten Szenarien die Methoden anwendbar sind und menschliche Verhaltensanalyse verwendet werden kann, um Erkenntnisse über den Prozess bzw. den Ablauf einer Studie zu erlangen.

Zunächst habe ich die interdisziplinäre Studie beschrieben und die zu behandelnden Hypothesen definiert. Durch den überwachten Klassifikationsansatz konnte ich darauf folgend Hypothese 8.1 (Menschen zeigen bereits vor dem Auftreten eines Fehlers ein Kopfverhalten, das den Fehler vorhersagbar macht.) bestätigen. Der Mensch zeigt bereits, bevor der Fehler auftritt, ein spezielles Verhalten, das durch maschinelle Lernansätze erkennbar ist. Als Fehler wird dabei der Moment angesehen, in dem das System diesen eigenständig erkennen kann. In diesem Fall ist es der Tod der Spielfigur. Die eigentliche Handlung, die zu dem Fehler führt, hat der Mensch bereits zuvor durch Tastendrucke durchgeführt. Diese Aktionen werden nicht betrachtet. Die Analyse in diesem Kapitel bezieht sich darauf, ob durch eine menschliche Verhaltensanalyse das System in die Lage versetzt werden kann, einen Fehler früher als sonst erkennen zu können. Dieses ist in dem beschriebenen Szenario der Fall.

Zur Untersuchung von Hypothese 8.2 (Anhand des Kopfverhaltens kann die steigende Komplexität und Schwierigkeit im Verlauf der Studie erkannt werden.) habe ich das vorgestellte unüberwachte Verfahren angewendet. Auch diese Hypothese kann bestätigt werden. Es wird einerseits die Mimik, andererseits die Anzahl der Versuche mit den beschriebenen Verfahren pro Level ausgewertet. In der unüberwachten Analyse kann deutlich die steigende Schwierigkeit der Level erkannt werden. Dieses löst sowohl eine andere Mimik, als auch eine andere Anzahl an Levelversuchen aus. Darüber hinaus können durch die Analyse der Mimik weitere Erkenntnisse zum Spielverlauf gewonnen werden. So wird im Gegensatz zu einer reinen Betrachtung der Versuche ein deutlicher Unterschied im gezeigten Verhalten im zweiten Level erkannt. Dieses kann wiederum durch den Spielablauf und der daraus folgenden erstmaligen Möglichkeit des Todes der Spielfigur erklärt werden.

Fazit

” *Wenn die anderen glauben, man ist am Ende, so muss man erst richtig anfangen.*

– **Konrad Adenauer**

Es ist eine große Herausforderung, objektive Bewertungen und konkrete, allumfassende Verbesserungsvorschläge zu prozessbasierten Systemen und zu Mensch-System Interaktionen zu erhalten. Dieses ist besonders in der nutzerorientierten Entwicklung der Fall, die zumeist mit subjektiven und fragebogengestützten Werkzeugen versucht, das Bauchgefühl und den inneren Zustand der Teilnehmenden zu erheben. Diese subjektiven Daten lassen zwar eine generelle Bewertung zu, schwere Störungen im Prozess überstrahlen aber oftmals leichte Schwierigkeiten, die darauffolgend durch die Versuchspersonen in der Evaluation nur selten genannt werden. Somit ist eine vollständige Bewertung und eine schnelle Verbesserung nur durch viele Iterationen im Entwicklungsprozess zu erlangen. Eine objektive Bewertung von prozessbasierten Systemen hingegen findet in der nutzerorientierten Entwicklung oft nicht oder nur unzureichend statt. Sie wird dabei nicht zuletzt durch die stetig steigende Anzahl von Assistenzsystemen in industriellen Umgebungen immer wichtiger, um eine gute Nutzerzufriedenheit und somit gute Arbeitsumgebungen auch bei der täglichen Arbeit mit technischen Hilfssystemen zu gewährleisten.

In dieser Arbeit habe ich deshalb objektive Verfahren vorgestellt, die eine Bewertung auf Basis einer multimodalen Analyse menschlichen Verhaltensweisen zulassen und den gesamten Prozess berücksichtigen. Ich habe mich dabei besonders auf die Analyse von Gesichtsausdrücken konzentriert, da das Gesicht eines der wichtigsten nonverbalen Kommunikationskanäle darstellt.

9.1 Zusammenfassung

Diese Arbeit zeigt, dass eine Analyse der Verhaltensweisen von Personen, die mit einem Assistenzsystem interagieren, zur Bewertung der einzelnen Prozessschritte verwendet werden kann (siehe Forschungsfrage 1). Zum anderen habe ich dargelegt, wie die entwickelten Verfahren, neben der Anwendung im Umfeld von Assistenzsystemen, auch in anderen Bereichen der Mensch-System Interaktion eingesetzt werden können (siehe Forschungsfrage 2). Für beide Teilaspekte habe ich darüber hinaus Methoden entwickelt, die in der Praxis verwendet werden können.

Dazu habe ich zunächst zur Motivation der entwickelten Verfahren den fachlichen Hintergrund dargestellt. Ich habe die Rolle und die Herausforderungen des Menschen in der industriellen Produktion der Zukunft erläutert, die Konzepte der nutzerorientierten Entwicklung eingeführt und die Gesichtsanalyse als Grundlage der menschlichen Zustands-erkennung vorgestellt. Darauf folgend habe ich eine von mir durchgeführte Umsetzung eines nutzerorientierten Entwicklungsprozesses für eine ganzheitliche Montageassistenzsystementwicklung in Kooperation mit einem lokalen Unternehmen beschrieben. Das daraus entstandene System habe ich als Grundlage für die Datenerhebung und zur Nutzung in den entwickelten Mustererkennungsverfahren verwendet. Es wurden Merkmalsgruppen abgeleitet, die ich für die menschliche Verhaltensanalyse verwendet habe. Weiterhin habe ich beschrieben, wie diese robust während der Interaktion von nutzenden Personen mit dem System detektiert werden können. Dabei wurde eine multimodale Kombination zwischen Kopfausdruck, Aufmerksamkeit, Systeminteraktion und Prozessschrittdauer verwendet. Um eine kontrollierbare Datenerhebung durchführen zu können, habe ich das von mir entwickelte Assistenzsystem für eine wissenschaftliche Auswertung angepasst. Ein wichtiger Schritt war dabei die Einführung von Störungen des Prozesses, um gezielt menschliche Verhaltensweisen in der Interaktion mit dem System auszulösen.

Auf Basis einer Datenerhebung mit in diesem Montageprozess unerfahrenen Versuchspersonen habe ich zunächst ein überwachtes Verfahren entwickelt, um Prozessstörungen zu erkennen. Durch eine statistische Analyse habe ich gezeigt, dass diese Störungen durch die Betrachtung von menschlichem Verhalten detektiert werden können. Des Weiteren habe ich mithilfe einer qualitativen Analyse dargestellt, dass falsche Klassifikationen der gelernten Verfahren anhand der Anweisungen in den jeweiligen Prozessschritten erklärbar sind. Es wurde deutlich, dass nicht nur manuell hinzugefügte Störungen erkannt, sondern darüber hinaus Fehler im Prozess aufgedeckt werden, die durch die durchgeführte

nutzerorientierte Entwicklung nicht aufgefallen waren. Daraus folgt, dass die von mir gemachte Klasseneinteilung nicht uneingeschränkt gültig war. Meine Annahme des optimalen Prozesses, der durch die nutzerorientierte Entwicklung entstanden sein sollte, musste verworfen werden, da auch in den nicht manuell angepassten Schritten Störungen auftraten. Darüber hinaus wurde auch in einigen als Störung modellierten Schritten scheinbar kein auffälliges Verhalten der Versuchspersonen erkannt. Auch die Modellierung von leichten Prozessstörungen zur Verunsicherung der Versuchspersonen konnten nicht die gewünschten Erfolge liefern, da es zumeist keinerlei erkennbare Verhaltensweisen der Versuchspersonen hervorgerufen hat.

Aufgrund der herausfordernden Modellierung der Störungserkennung als binäres und überwachttes Entscheidungsproblem habe ich unüberwachte Verfahren zur Erkennung von Abweichungen im menschlichen Verhalten bei der Erfüllung prozessbasierter Aufgabenschritte betrachtet. Dabei habe ich Verfahren entwickelt, um ein Modell des durchschnittlichen menschlichen Verhaltens während der jeweiligen Prozessschrittdurchführung zu lernen. Mithilfe dieses Modells konnte ich einzelne Schritte vergleichen und diesen einen Gütewert in Form einer Distanz zuordnen. Ich konnte darüber hinaus durch eine qualitative Analyse der Prozessschritte mit den schlechtesten Gütewerten zeigen, dass diese Schritte mit Schwierigkeiten im Prozessablauf in Verbindung gebracht werden können. Somit ist eine Erkennung, Analyse und Verbesserung von nicht optimalen Aufgabenschritten auf dieser objektiven Basis möglich. Zur intuitiven Visualisierung und einfachen Interpretation der Ergebnisse der unüberwachten Methoden habe ich eine Heatmap gewählt.

Abschließend konnte ich zeigen, dass sowohl die entwickelten Konzepte und Verfahren der überwachten Klassifikation als auch die der unüberwachten Analyse auf andere Bereiche der Mensch-System Interaktion übertragbar sind. Dabei habe ich beide auf ein Lernszenario angewendet und konnte menschliche Fehler anhand einer menschlichen Verhaltensanalyse vorhersagen. Außerdem habe ich dargestellt, dass auch in diesem Szenario durch unüberwachte Analyseverfahren Erkenntnisse über den Ablauf der Studie und damit der Mensch-System Interaktion extrahiert werden können. Durch eine objektive Verhaltensanalyse konnte ich auf einzelne Schritte eingehen und diese interpretieren. Somit können die entwickelten Verfahren auch als Werkzeug in der Analyse wissenschaftlicher Studien wie beispielsweise zur Grundlagenforschung genutzt werden.

Ich konnte bei den in dieser Arbeit vorgestellten Verfahren der multimodalen menschlichen Verhaltensanalyse sowohl den praktischen Nutzen in der Entwicklung von prozessbasierten Assistenzsystemen als auch die Relevanz als Werkzeug für Assistenzsystementwickler und

-optimierer zeigen. Durch einen vollständigen Wechsel des Anwendungsbereichs konnte ich darüber hinaus die Generalisierbarkeit der Methoden nachweisen.

9.2 Ausblick

Ich sehe besonders bei dem entwickelten unüberwachten Ansatz ein hohes Potential für einen praktischen Einsatz als Werkzeug in der Entwicklung von menschenorientierten Systemen. Dieses gilt sowohl in der Analyse von prozessbasierten Assistenzsystemen als auch in der Analyse von Mensch-Maschine Interaktionen im Allgemeinen. Dabei kann der Ansatz die Basis einer Vielzahl von Anwendungen sein. Es kann beispielsweise auf Grundlage der Heatmap-Visualisierung ein intuitives Werkzeug für den Prozess- und Assistenzsystementwickler geschaffen werden, welches Assistenzsysteme objektiv bewertet und Potentiale zur Verbesserung der zugrundeliegenden Prozesse aufdeckt. Eine hierarchische Aufbereitung und Verarbeitung mit unterschiedlichen Dimensionalitäten wäre dabei am sinnvollsten, um sowohl einzelne Dimensionen als auch eine multimodale Auswertung in der Analyse berücksichtigen zu können.

Ich vermute darüber hinaus, dass ein einzelnes Modell zur Abbildung des normalen Nutzungsverhaltens in umfangreichen Szenarien nicht ausreicht. Bei zusätzlicher Verwendung von Informationen über die Art der einzelnen Prozessschritte (beispielsweise Komplexität) könnte eine weitaus detailliertere Auswertung erstellt werden.

Auch die verwendeten Merkmalsgruppen und definierten Merkmale bieten weitere Möglichkeiten der Verbesserung und Erweiterung. So verwendete ich Merkmale, die in der Lage sind, zeitliche Abhängigkeiten basal zu modellieren. Im Lernszenario habe ich bereits weitere zeitliche Merkmale berücksichtigt, gerade bei der zeitlichen Betrachtung einzelner Merkmale und dem Vergleichen von Zeiträumen unterschiedlicher Länge sind aber dennoch deutlich komplexere Methoden vorhanden, die detailliertere zeitliche Variationen von Merkmalen modellieren und berücksichtigen können. Eine Verwendung solcher Modelle wäre wünschenswert, um diese mit den von mir verwendeten Merkmalen vergleichen zu können und so mitunter eine bessere Erkennungsrate zu erzielen und weitere Potentiale im Prozessablauf aufdecken zu können. Auch eine Anwendung der Merkmale aus dem Lernszenario auf das Montageszenario und einer Verwendung der beschriebenen kurzen Kontrollblicke zwischen System und Werkstück wäre sinnvoll. Des Weiteren wäre eine Anwendung der im Montageszenario gelernten konkreten Modelle auf das Lernszenario

eine zusätzliche Möglichkeit in Zukunft die Generalisierbarkeit der Verfahren und nicht nur der reinen Methodik untermauern zu können.

Außerdem habe ich mich dafür entschieden, sämtliche Gesichtsausdrücke gleichwertig zu behandeln und keine Unterschiede zwischen ihnen zu machen. Aus der Psychologie ist jedoch bekannt, dass bestimmte Gesichtsausdrücke mit bestimmten inneren menschlichen Zuständen und Emotionen in Verbindung gebracht werden können. Für zukünftige Entwicklungen kann eine Nutzung dieser Informationen und eine Betrachtung von unterschiedlichen Arten der Gesichtsmimik weitere und genauere Erkenntnisse über den Prozess und die Mensch-System Interaktion generieren. Diese können wiederum gewinnbringend in den von mir entwickelten Lernverfahren verwendet werden.

Sowohl die vorgestellten überwachten als auch die unüberwachten Methoden können für eine automatische Systemanpassung verwendet und permanent zur Unterstützung der arbeitenden Personen eingesetzt werden. Besonders wenn davon ausgegangen werden kann, dass der Prozess optimal ist und dennoch individuelle Schwierigkeiten bei der Prozessdurchführung auftreten, könnte ein System autonom auf diese erkannten Schwierigkeiten reagieren. Es könnte außerdem mögliche Lösungsvorschläge präsentieren bzw. weitere Informationen zur Aufgabenlösung der individuell nutzenden Person zur Verfügung stellen. Somit wäre ein Assistenzsystem in der Lage, sich an diese anzupassen, je nachdem welches Verhalten sie zeigt. Wie genau eine permanenten Unterstützung konkret aussehen kann, ohne die Arbeitssituation und -bedingungen durch das technische System und die permanente Beobachtung zu verschlechtern, muss in zukünftigen Forschungsprojekten ermittelt werden. Ein solches System müsste in Langzeitstudien detailliert evaluiert werden und wirft darüber hinaus ethische Fragestellungen im Bereich des Datenschutzes und der permanenten Überwachung und Kontrolle der arbeitenden Personen auf.

Auch die Integration der vorgestellten Analysen in die Datenvielfalt einer vernetzten Produktion und das Wechselspiel zwischen automatischen Analysen von maschinellen und menschlichen Verhaltens in einer Fabrik sollte in zukünftigen Projekten in Hinblick auf die Arbeitsbedingung in der Produktion der Zukunft betrachtet werden.

Zusammenfassend leistet diese Arbeit einen praxisrelevanten Beitrag in der Mensch-System Interaktionsforschung. Die hier vorgestellten Verfahren können die nutzerorientierte Entwicklung beschleunigen und als Grundlage dienen, um prozessbasierte Systeme objektiv zu bewerten und darauf aufbauend diese Systeme zu verbessern sowie Erkenntnisse über die zugrundeliegenden Verhaltensprozesse zu erlangen.

Zusätzliches Material der Datenerhebung in der Montage

A.1 Studienmaterial

A.1.1 Stellungnahme der Ethik-Kommission

Universität Bielefeld Ethik-Kommission

Ethik-Kommission der Universität Bielefeld
Postfach 10 01 31 | D-33501 Bielefeld

Der Vorsitzende

Geschäftsstelle:
Fatma Akkaya-Willis
Raum: T5-239
Tel.: 0521 106-4436
ethikkommission@uni-bielefeld.de
Az.: 1266

Bielefeld, 06.09.2017
Seite 1 von 1

**Stellungnahme der Ethik-Kommission der Universität Bielefeld zu Antrag
Nr. 2017- 174 vom 06.09.2017**

Kurzbezeichnung der Studie: „**Menschliches Verhalten in Montageumgebungen**“

Hauptansprechpartner: Torben Töniges
Betreuerin: Britta Wrede

Die Ethik-Kommission der Universität Bielefeld hat den Antrag nach den ethischen Richtlinien der Deutschen Gesellschaft für Psychologie e.V. und des Berufsverbands Deutscher Psychologinnen und Psychologen e.V. begutachtet.

Auf der Grundlage der eingereichten Unterlagen hält die Ethik-Kommission der Universität Bielefeld die Durchführung der Studie in der beschriebenen Form für ethisch unbedenklich.

Für die Ethik-Kommission

Fridtjof Nußbeck
Prof. Dr. Fridtjof W. Nußbeck

A.1.2 Einverständniserklärung

Das Technologie-Netzwerk:
Intelligente Technische Systeme
OstWestfalenLippe



Einverständniserklärung für Teilnehmer/innen

Ihre Einwilligung und die Teilnahme an der Evaluation sind freiwillig. Sie können Ihre Einwilligung jederzeit mit Wirkung für die Zukunft widerrufen und die Löschung bzw. Vernichtung ihrer Daten und der Video- und Audioaufnahmen verlangen. Durch eine Verweigerung oder einen Widerruf der Einwilligungserklärung entstehen Ihnen keine Nachteile.

Mit Ihrer Unterschrift bestätigen Sie, dass Sie über den Ablauf in verständlicher Form aufgeklärt wurden und damit einverstanden sind.

Darüber hinaus erklären Sie sich mit folgenden Punkten einverstanden:

– Punkte, denen Sie nicht zustimmen, bitte streichen –

1. An der Teilnahme der Evaluation mit der oben beschriebenen Aufnahme, Speicherung und Verarbeitung der Daten im Rahmen dieses Transferprojektes (itsowl-TT-IMEMO: Hanning Elektro-Werke GmbH & Co. KG & Research Institute for Cognition and Robotics, CoR-Lab, Universität Bielefeld).
2. Mit der Weitergabe der Video- und Audiodaten innerhalb des Transferprojektes und der beteiligten Arbeitsgruppen. Ohne Ihre Einwilligung erfolgt keine Weitergabe der Daten an Dritte.
3. Verwendung der Video- und Audiodaten für Präsentationszwecke mit unkenntlich gemachten Teilnehmern.
4. Mit der Verwendung Ihrer ausgefüllten, anonymisierten Fragebögen innerhalb des Transferprojektes und der beteiligten Arbeitsgruppen.

Vielen Dank für Ihre Teilnahme an der Evaluation.

Persönliche Angaben der/s Teilnehmerin/s

Name: _____

Vorname: _____

Telefon oder andere Kontaktinformationen:

Datum, Unterschrift der/s Teilnehmerin/s:

A.1.3 System Usability Score (SUS) Fragebogen

Allgemeiner Abschlussfragebogen

Datum:

Uhrzeit:

Ich kann mir sehr gut vorstellen, das System regelmäßig zu nutzen.

Stimme ich gar nicht zu

Stimme ich voll zu

Ich empfinde das System als unnötig komplex.

Stimme ich gar nicht zu

Stimme ich voll zu

Ich empfinde das System als einfach zu nutzen.

Stimme ich gar nicht zu

Stimme ich voll zu

Ich denke, dass ich technischen Support brauchen würde, um das System zu nutzen.

Stimme ich gar nicht zu

Stimme ich voll zu

Ich finde, dass die verschiedenen Funktionen des Systems gut integriert sind.

Stimme ich gar nicht zu

Stimme ich voll zu

Ich finde, dass es im System zu viele Inkonsistenzen gibt.

Stimme ich gar nicht zu

Stimme ich voll zu

Ich kann mir vorstellen, dass die meisten Leute das System schnell zu beherrschen lernen.

Stimme ich gar nicht zu

Stimme ich voll zu

Ich empfinde die Bedienung als sehr umständlich.

Stimme ich gar nicht zu

Stimme ich voll zu

Ich habe mich bei der Nutzung des Systems sehr sicher gefühlt.

Stimme ich gar nicht zu

Stimme ich voll zu

Ich musste eine Menge Dinge lernen, bevor ich mit dem System arbeiten konnte.

Stimme ich gar nicht zu

Stimme ich voll zu

A.1.5 Persönlichkeitsfragebogen

Persönlichkeitsfragebogen auf Basis von Rammstedt u. a. [Ram+12]:

Montageassistentz

Herzlich Willkommen zum abschließenden Fragebogen der Montageassistentzstudie.

Diese Umfrage enthält 6 Fragen.

Persönliche Informationen

[]Bitte geben Sie Ihre eMail-Adresse an: *

Bitte geben Sie Ihre Antwort hier ein:

[]Bitte wählen Sie Ihr Geschlecht: *

Bitte wählen Sie nur eine der folgenden Antworten aus:

- weiblich
 männlich

[]Bitte geben Sie Ihr Alter an: *

In dieses Feld dürfen nur Zahlen eingegeben werden.

Bitte geben Sie Ihre Antwort hier ein:

[]Bitte geben Sie an, wieviel Erfahrung Sie haben mit: *

Bitte wählen Sie die zutreffende Antwort für jeden Punkt aus:

	1 (keine)	2	3	4	5	6 (sehr viel)
Nutzung von Computern	<input type="radio"/>					
Nutzung von Montageassistentzsystemen	<input type="radio"/>					
Nutzung von Handy/Tablet	<input type="radio"/>					

[] Inwieweit treffen die folgenden Aussagen auf Sie zu? *

Bitte wählen Sie die zutreffende Antwort für jeden Punkt aus:

	trifft überhaupt nicht zu	trifft eher nicht zu	weder noch	eher zutreffend	trifft voll und ganz zu
Ich bin eher zurückhaltend, reserviert.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ich schenke anderen leicht Vertrauen, glaube an das Gute im Menschen.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ich bin bequem, neige zur Faulheit.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ich bin entspannt, lasse mich durch Stress nicht aus der Ruhe bringen.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ich habe nur wenig künstlerisches Interesse.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ich gehe aus mir heraus, bin gesellig.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ich neige dazu, andere zu kritisieren.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ich erledige Aufgaben gründlich.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ich werde leicht nervös und unsicher.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ich habe eine aktive Vorstellungskraft, bin fantasievoll.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

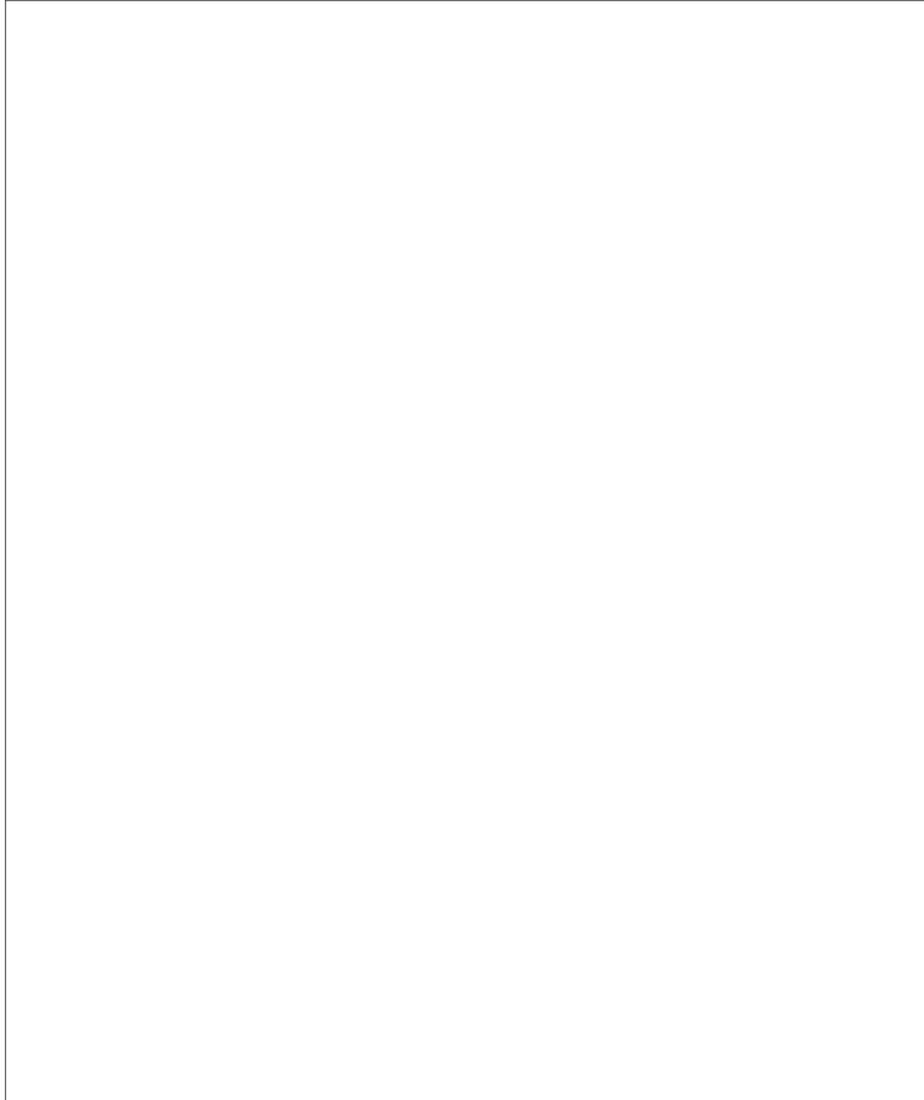
Kommentare

[]

Haben Sie weitere Kommentare zur Studienteilnahme?

Haben Sie Verbesserungsvorschläge für das Assistenzsystem? Was ist Ihnen besonders negativ / positiv aufgefallen?

Bitte geben Sie Ihre Antwort hier ein:



A.2 ROC-Kurven unterschiedlicher Merkmalskombinationen

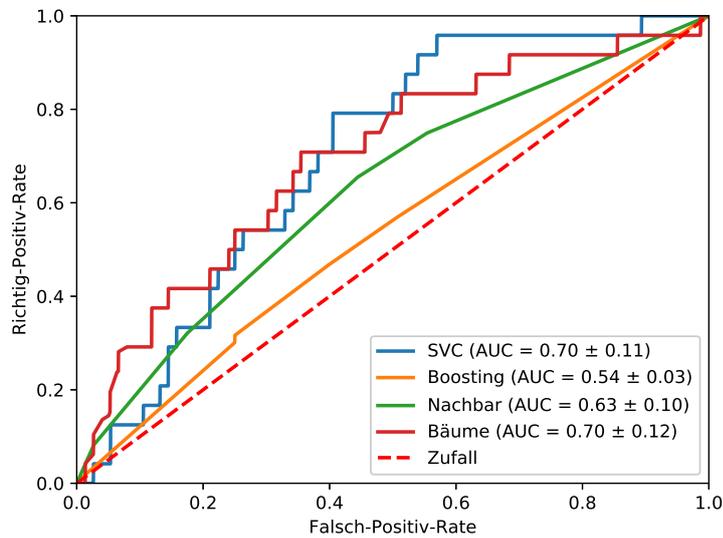


Abbildung A.1.: ROC-Kurven der verwendeten Klassifikatoren bei Nutzung der A-K Merkmalsgruppe

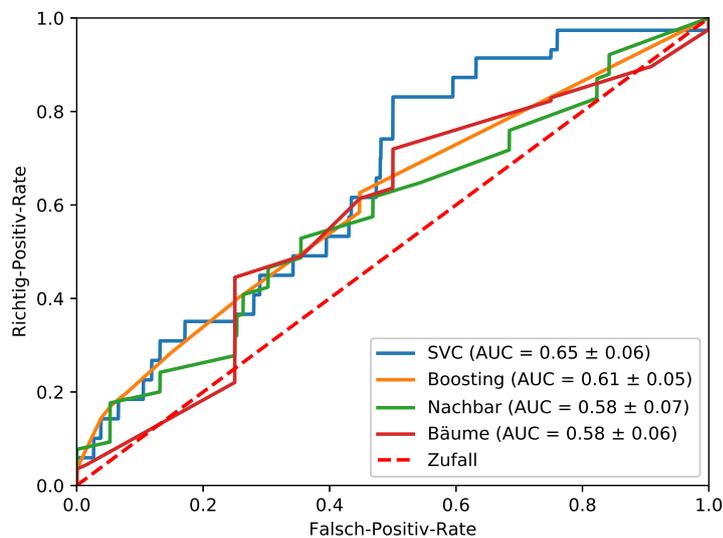


Abbildung A.2.: ROC-Kurven der verwendeten Klassifikatoren bei Nutzung der A-R Merkmalsgruppe

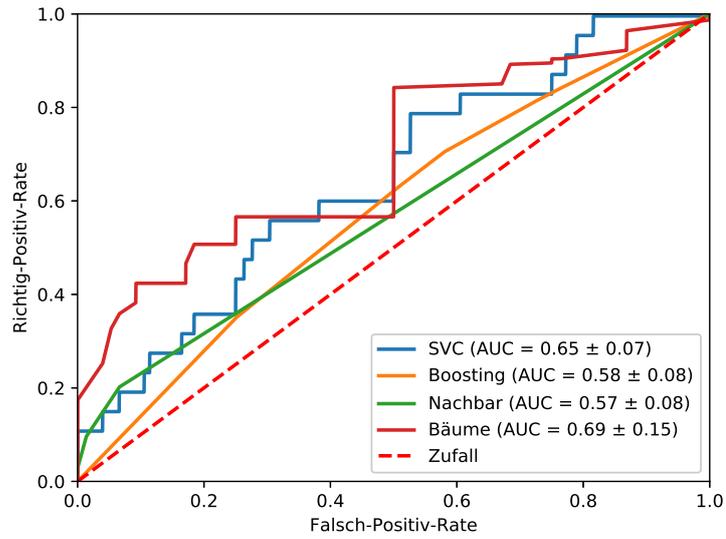


Abbildung A.3.: ROC-Kurven der verwendeten Klassifikatoren bei Nutzung der As-K Merkmalsgruppe

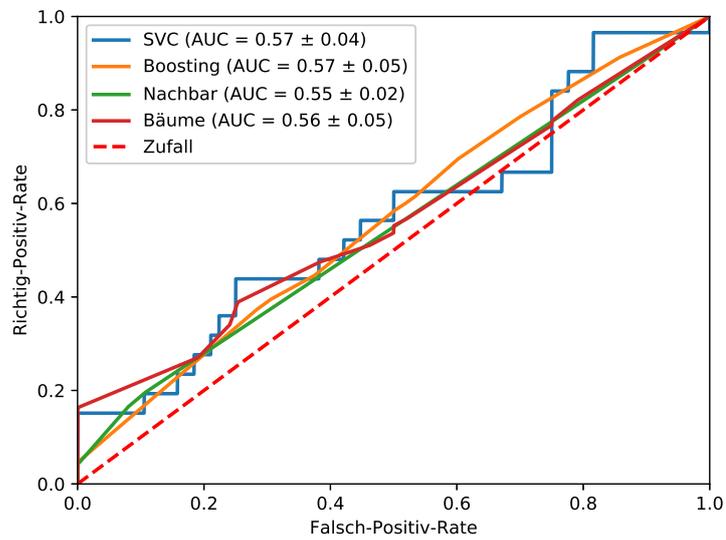


Abbildung A.4.: ROC-Kurven der verwendeten Klassifikatoren bei Nutzung der As-R Merkmalsgruppe

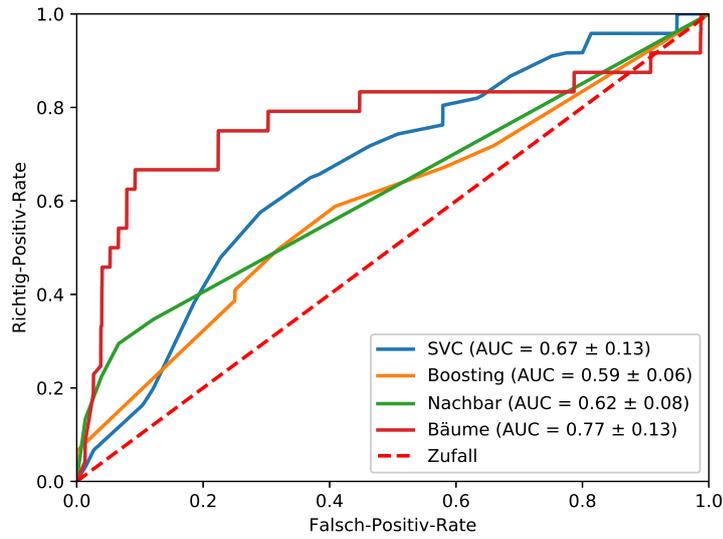


Abbildung A.5.: ROC-Kurven der verwendeten Klassifikatoren bei Nutzung der A-KZ Merkmalsgruppen

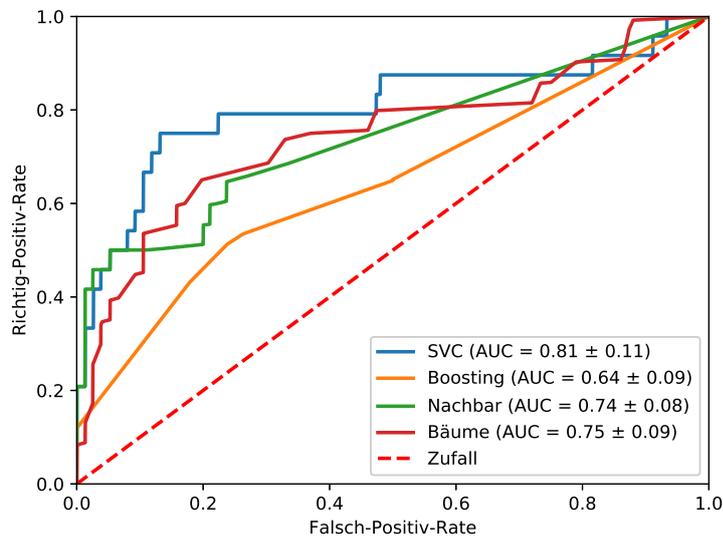


Abbildung A.6.: ROC-Kurven der verwendeten Klassifikatoren bei Nutzung der A-RZ Merkmalsgruppen

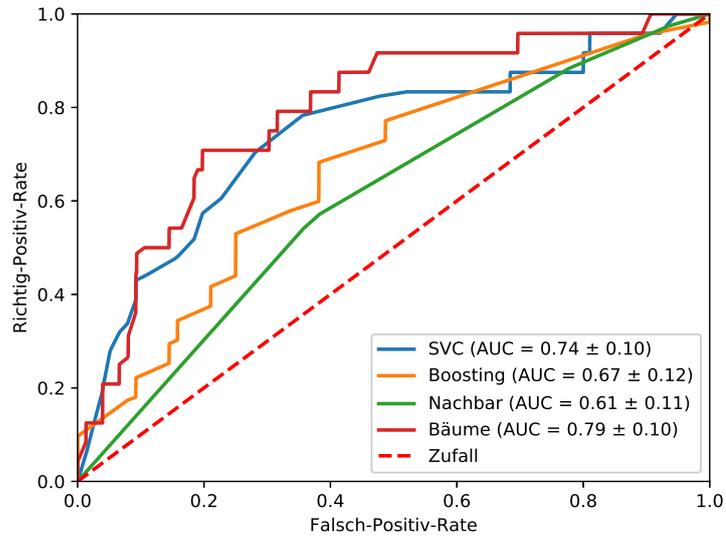


Abbildung A.7.: ROC-Kurven der verwendeten Klassifikatoren bei Nutzung der A-KSB Merkmalsgruppen

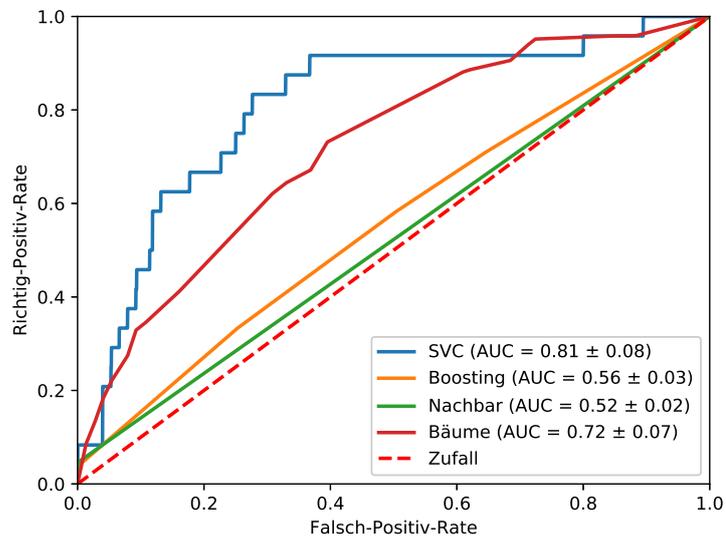


Abbildung A.8.: ROC-Kurven der verwendeten Klassifikatoren bei Nutzung der A-RSB Merkmalsgruppen

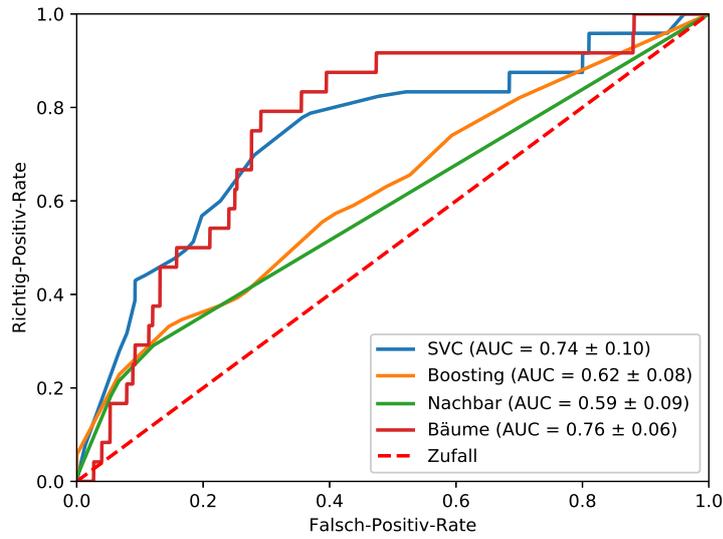


Abbildung A.9.: ROC-Kurven der verwendeten Klassifikatoren bei Nutzung der A-KSBZ Merkmalsgruppen

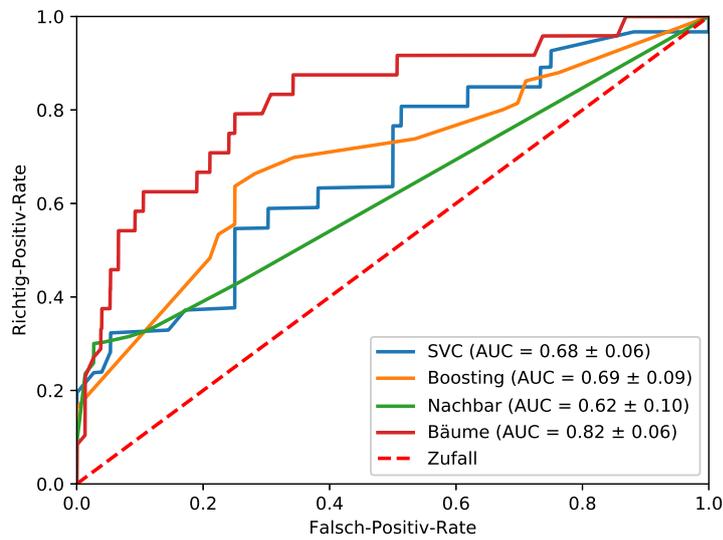


Abbildung A.10.: ROC-Kurven der verwendeten Klassifikatoren bei Nutzung der A-RSBZ Merkmalsgruppen

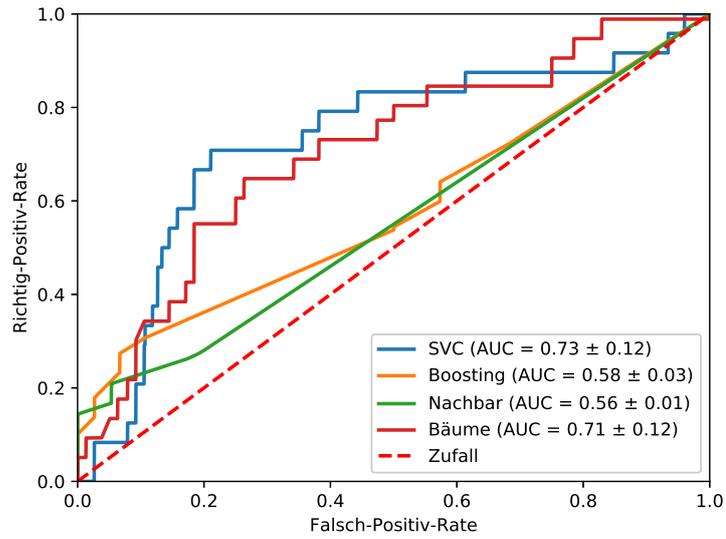


Abbildung A.11.: ROC-Kurven der verwendeten Klassifikatoren bei Nutzung der A-KAS-KSBZ Merkmalsgruppen

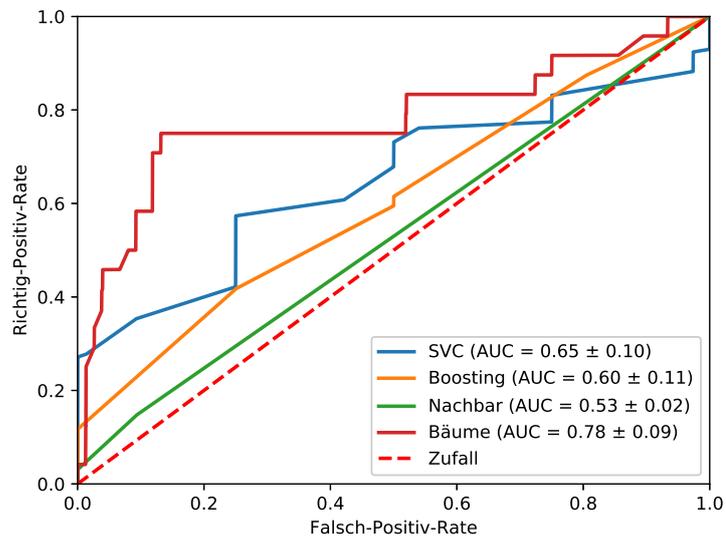


Abbildung A.12.: ROC-Kurven der verwendeten Klassifikatoren bei Nutzung der A-RAS-RSBZ Merkmalsgruppen

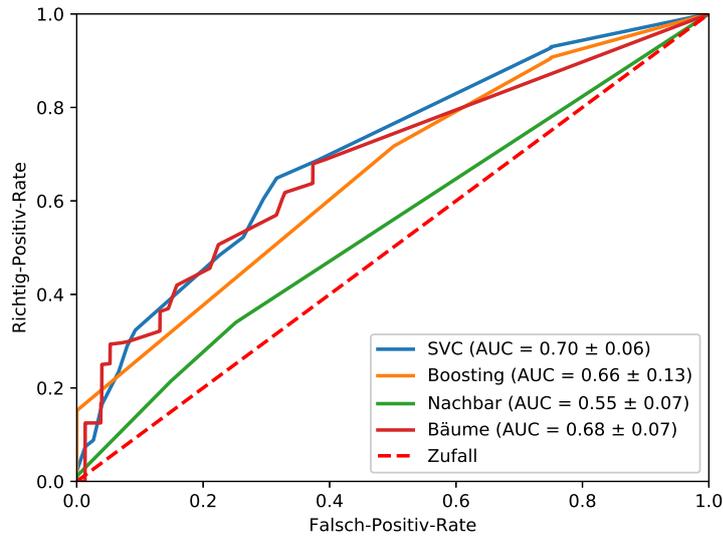


Abbildung A.13.: ROC-Kurven der verwendeten Klassifikatoren bei Nutzung der S Merkmalsgruppe

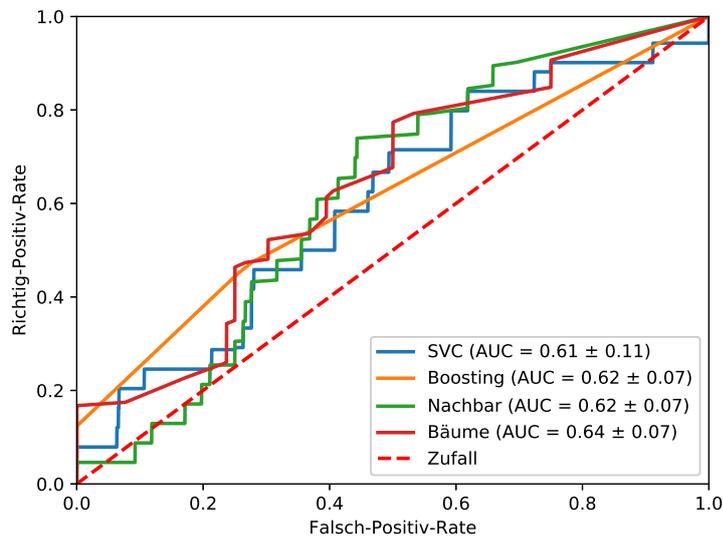


Abbildung A.14.: ROC-Kurven der verwendeten Klassifikatoren bei Nutzung der B Merkmalsgruppe

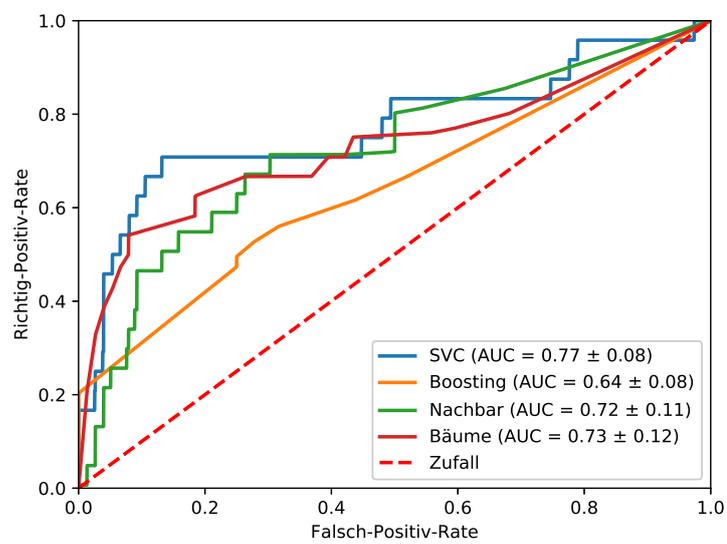


Abbildung A.15.: ROC-Kurven der verwendeten Klassifikatoren bei Nutzung der Z Merkmalsgruppe

Zusätzliches Material der Studie zum Methodentransfer

B.1 Studienmaterial

B.1.1 Stellungnahme der Ethik-Kommission

Universität Bielefeld Ethik-Kommission

Ethik-Kommission der Universität Bielefeld
Postfach 10 01 31 | D-33501 Bielefeld

Der Vorsitzende

Geschäftsstelle:
Fatma Akkaya-Willis
Raum: T5-239
Tel.: 0521 106-4438
ethikkommission@uni-bielefeld.de
Az.: 1266

Bielefeld, 10.05.2017
Seite 1 von 1

**Stellungnahme der Ethik-Kommission der Universität Bielefeld zu Antrag
Nr. 2017 – 096 vom 10.05.2017**

Kurzbezeichnung der Studie: **“Blick und Reaktion beim Computerspiel“**

Hauptansprechpartner: Torben Töniges
Betreuerin: Britta Wrede

Die Ethik-Kommission der Universität Bielefeld hat den Antrag nach den ethischen Richtlinien der Deutschen Gesellschaft für Psychologie e.V. und des Berufsverbands Deutscher Psychologinnen und Psychologen e.V. begutachtet.

Auf der Grundlage der eingereichten Unterlagen hält die Ethik-Kommission der Universität Bielefeld die Durchführung der Studie in der beschriebenen Form für ethisch unbedenklich.

Für die Ethik-Kommission


Prof. Dr. Gerd Bohner
Vorsitzender

B.1.2 Studienbeschreibung

Informationen zur Studie

Im nachfolgenden Experiment geht es um die Erfassung von Reaktionen beim Videospiel sowie die Aufzeichnung Ihrer Augenbewegungen und Ihres Gesichtsausdrucks. Sie spielen ein Computerspiel für ca. 5 Minuten mit steigender Schwierigkeit. Die Studie ist als Wettbewerb ausgelegt, die besten drei TeilnehmerInnen erhalten einen Preis (Einkaufsgutschein). Sie bedienen das Computerspiel mit den Pfeiltasten des Computers. Damit Sie das Spiel kennen lernen, sehen Sie zu Beginn ein kurzes Video.

Während des Experiments werden Ihre Blickbewegungen und Ihr Gesichtsausdruck mit Hilfe einer Kamera aufgezeichnet. Aus dem daraus entstehenden Video werden die benötigten Daten berechnet. Die daraus resultierenden Forschungsdaten werden in Bild 1 dargestellt. Um eine möglichst genaue Aufzeichnung zu ermöglichen, ist es wichtig, dass Sie so bequem wie möglich sitzen, so dass Sie Ihre Körperhaltung für die Dauer des Experiments beibehalten können. Damit das System Ihre Augen adäquat erkennt, führen wir vor Beginn eine Kalibrierung durch. Eine genaue Anleitung dazu erhalten Sie gleich auf dem Bildschirm.

Nach dem Computerspiel folgt noch ein kurzer Fragebogen, bei dem jedoch keine Kamaraufzeichnung stattfindet.

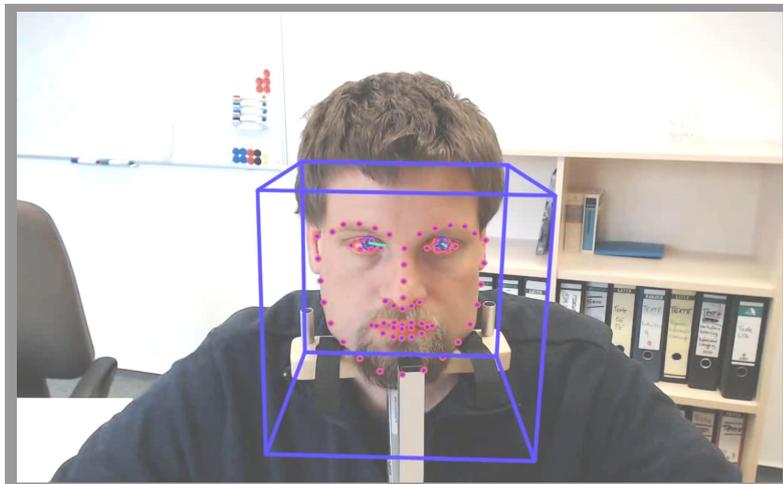


Bild 1: Hier sehen Sie die aus dem Gesichtsausdruck berechneten Forschungsdaten.

B.1.3 Einverständniserklärung

Einverständniserklärung

Hinweise zum Datenschutz

Ihre Angaben und Daten werden von den Projektpartnern der Universität Paderborn (Prof. Dr. Christian Harteis, Dr. Dagmar Festner, Christoph Fischer) und der Universität Bielefeld (Prof. Dr. Britta Wrede, Torben Töniges) streng vertraulich behandelt. Die Ergebnisse werden nur anonymisiert weiterverwendet. Die Teilnahme an der Studie ist freiwillig. Es entstehen Ihnen keine Nachteile, wenn Sie nicht teilnehmen. Sie haben auch die Möglichkeit, einzelne Fragen nicht zu beantworten. Bei allen Veröffentlichungen (z.B. Berichten, Präsentationen, wissenschaftlichen Veröffentlichungen) werden die Daten so aufbereitet, dass kein Rückschluss auf Ihre Person möglich ist.

Einverständniserklärung

Ich habe die Hinweise zum Datenschutz gelesen und bin mit der dort beschriebenen Vorgehensweise zum Datenschutz einverstanden.

Ich willige ein, dass die von mir gemachten Angaben sowie die erfassten Augenbewegungen und das Gesichtsabbild gespeichert und für wissenschaftliche Zwecke verwendet werden dürfen.

Ich bin darüber informiert, dass die Aufzeichnung und Auswertung aller Daten zu keinem Zeitpunkt zusammen mit meinem Namen gespeichert werden. Die aufgezeichneten Daten werden nur von Personen ausgewertet, die auf das Datengeheimnis verpflichtet wurden und die keine vertraulichen und personenbezogenen Informationen an Dritte weitergeben.

Ich wurde darüber informiert, dass ich keine der gestellten Fragen beantworten muss und dass die Ergebnisse nur in einer Form präsentiert werden, die einen Rückschluss auf mich nicht zulässt.

Ich willige unter den oben genannten Bedingungen ein, an der Studie teilzunehmen.

Paderborn

Ort, Datum

Unterschrift

Zusätzliche Einverständniserklärung (optional)

Zur Veranschaulichung der Forschung durch Bildaufnahmen benötigen wir von Ihnen eine zusätzliche Einverständniserklärung. Um die Anonymität zu gewährleisten, wird von Ihnen nur eine Identifikationsnummer gespeichert. Diese Nummer wird nur im Falle Ihres nachträglichen Widerrufs zu Ihrer Einverständniserklärung mit den Forschungsdaten zusammengeführt. In allen Publikationen wird kein anderer Hinweis als Ihr Bild auf Sie persönlich verweisen.

Hiermit willige ich zur Veröffentlichung meines Bildes in folgenden Formaten ein:

Konferenzen und Tagungen:	<i>Video</i> <input type="checkbox"/>	<i>Bild</i> <input type="checkbox"/>
Publikationen:	<i>Video</i> <input type="checkbox"/>	<i>Bild</i> <input type="checkbox"/>

Hiermit erkläre ich mich damit einverstanden, dass die oben bezeichneten Foto- bzw. Videoaufnahmen von mir angefertigt, zum genannten Zweck eingesetzt und wie aufgeführt veröffentlicht werden können. Aus der Zustimmung zur Veröffentlichung leite ich keine Rechte (z. B. Entgelt) ab. Diese Einverständniserklärung ist gegenüber dem Veranlasser jederzeit mit Wirkung für die Zukunft widerrufbar. Im Falle des Widerrufs werden die Aufnahmen ab dem Zeitpunkt des Zugangs des Widerrufs nicht mehr für Veröffentlichungen eingesetzt. Bereits veröffentlichte Aufnahmen bleiben unangetastet.

Paderborn

Ort, Datum

Name, Vorname

Unterschrift

B.1.4 Persönlichkeitsfragebogen

Persönlichkeitsfragebogen auf Basis von Rammstedt u. a. [Ram+12]:

Blick und Reaktion beim Computerspiel

Fragebogen

Nach dem Experiment möchten wir Sie noch um das Ausfüllen dieses Fragebogens bitten.

Kreuzen Sie bei den folgenden 10 Begriffspaaren Ihrem Empfinden entsprechend intuitiv an.

Das Spielen des Computerspiels war für mich:

Wichtig	<input type="checkbox"/>	Unwichtig						
Interessant	<input type="checkbox"/>	Uninteressant						
Relevant	<input type="checkbox"/>	Irrelevant						
Aufregend	<input type="checkbox"/>	Nicht aufregend						
Von hoher Bedeutung	<input type="checkbox"/>	Bedeutet mir nichts						
Ausdrucksvoll	<input type="checkbox"/>	Ausdruckslos						
Spannend	<input type="checkbox"/>	Langweilig						
Von hohem Stellenwert	<input type="checkbox"/>	Von geringem Stellenwert						
Involvierend	<input type="checkbox"/>	Nicht Involvierend						
Anregend	<input type="checkbox"/>	Nicht Anregend						

Inwieweit treffen die folgenden Aussagen auf Sie zu?

	<i>Stimme gar nicht zu</i>	<i>Stimme eher nicht zu</i>	<i>Stimme teils/teils zu</i>	<i>Stimme eher zu</i>	<i>Stimme voll zu</i>
Ich bin eher zurückhaltend, reserviert.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Ich schenke anderen leicht Vertrauen, glaube an das Gute im Menschen.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Ich bin bequem, neige zur Faulheit.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Ich bin entspannt, lasse mich durch Stress nicht aus der Ruhe bringen.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Ich habe nur wenig künstlerisches Interesse.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Ich gehe aus mir heraus, bin gesellig.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Ich neige dazu, andere zu kritisieren.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Ich erledige Aufgaben gründlich.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Ich werde leicht nervös und unsicher.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Ich habe eine aktive Vorstellungskraft, bin fantasievoll.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Literaturverzeichnis

- [AMP04] Chadia Abras, Diane Maloney-Krichmar und Jenny Preece. “User-centered design”. In: *Bainbridge, W. Encyclopedia of Human-Computer Interaction. Thousand Oaks: Sage Publications* 37.4 (2004), S. 445–456 (siehe S. 10).
- [ACB02] Steve Adolph, Alistair Cockburn und Paul Bramble. *Patterns for effective use cases*. Addison-Wesley Longman Publishing Co., Inc, 2002 (siehe S. 11).
- [Agu04] Ruth Sara Aguilar-Savén. “Business process modelling: Review and framework”. In: *International Journal of Production Economics* 90.2 (2004), S. 129–149 (siehe S. 3).
- [Alt+17] Katharina Altemeier, Michael Bansmann, Oliver Ditrich, Roman Dumitrescu und Wolfgang Nettelstroth. “Auf dem Weg zu Industrie 4.0: Gestaltung Digitalisierter Arbeitswelten”. *it’s OWL Clustermanagement GmbH*, 2017 (siehe S. 1, 5).
- [Ann02] John Annett. “Subjective rating scales: science or art?” In: *Ergonomics* 45.14 (2002), S. 966–987 (siehe S. 12).
- [Ash+09] Ahmed Bilal Ashraf, Simon Lucey, Jeffrey F. Cohn u. a. “The Painful Face - Pain Expression Recognition Using Active Appearance Models”. In: *Image and Vision Computing* 27.12 (2009), S. 1788–1796 (siehe S. 20).
- [BG11] Radhakisan Baheti und Helen Gill. “Cyber-physical systems”. In: *The impact of control technology* 12 (2011), S. 161–166 (siehe S. 6).
- [Bai83] Lisanne Bainbridge. “Ironies of automation”. In: *Analysis, Design and Evaluation of Man-Machine Systems*. Hrsg. von G. Johannsen und J. E. Rijnsdorp. Elsevier, 1983, S. 129–135 (siehe S. 7).
- [BMR15] Tadas Baltrusaitis, Marwa Mahmoud und Peter Robinson. “Cross-dataset learning and person-specific normalisation for automatic Action Unit detection”. In: *2015 11th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG)*. Piscataway, NJ: IEEE, 2015, S. 1–6 (siehe S. 45, 46).

- [BRM13] Tadas Baltrusaitis, Peter Robinson und Louis-Philippe Morency. “Constrained Local Neural Fields for Robust Facial Landmark Detection in the Wild”. In: *The IEEE International Conference on Computer Vision (ICCV) Workshops*. 2013 (siehe S. 43).
- [BRM16] Tadas Baltrusaitis, Peter Robinson und Louis-Philippe Morency. “OpenFace: An open source facial behavior analysis toolkit”. In: *2016 IEEE Winter Conference on Applications of Computer Vision (WACV)*. Piscataway, NJ: IEEE, 2016, S. 1–10 (siehe S. 40–47).
- [BR11] Ntombikayise Banda und Peter Robinson. “Multimodal Affect Recognition in Intelligent Tutoring Systems”. In: *Affective Computing and Intelligent Interaction* (2011), S. 200–207 (siehe S. 21).
- [Ber+06] L. M. Bergasa, J. Nuevo, M. A. Sotelo, R. Barea und M. E. Lopez. “Real-Time System for Monitoring Driver Vigilance”. In: *IEEE Transactions on Intelligent Transportation Systems* 7.1 (2006), S. 63–77 (siehe S. 20).
- [Bor+14] Gianluca Borghini, Laura Astolfi, Giovanni Vecchiato, Donatella Mattia und Fabio Babiloni. “Measuring neurophysiological signals in aircraft pilots and car drivers for the assessment of mental workload, fatigue and drowsiness”. In: *Neuroscience and biobehavioral reviews* 44 (2014), S. 58–75 (siehe S. 15).
- [BH15] Alfons Botthof und Ernst Andreas Hartmann. *Zukunft der Arbeit in Industrie 4.0*. Berlin und Heidelberg: Springer Berlin Heidelberg und Springer Vieweg, 2015 (siehe S. 5).
- [Bra97] Andrew P. Bradley. “The use of the area under the ROC curve in the evaluation of machine learning algorithms”. In: *Pattern Recognition* 30.7 (1997), S. 1145–1159 (siehe S. 83).
- [Bre01] Leo Breiman. “Random Forests”. In: *Machine Learning* 45.1 (2001), S. 5–32 (siehe S. 82).
- [Bro13] John Brooke. “SUS: A Retrospective”. In: *J. Usability Studies* 8.2 (2013), S. 29–40 (siehe S. 37, 73).
- [Bro+96] John Brooke u. a. “SUS-A quick and dirty usability scale”. In: *Usability evaluation in industry* 189.194 (1996), S. 4–7 (siehe S. 13).
- [Bun17] Bundesministerium für Arbeit und Soziales. “Weissbuch Arbeiten 4.0”. 2017 (siehe S. 1).
- [Cai07] Brad Cain. *A Review of the Mental Workload Literature*. Hrsg. von Defence Research And Development Toronto. 2007 (siehe S. 12).

- [CW84] J. G. Casali und W. W. Wierwille. “On the measurement of pilot perceptual workload: a comparison of assessment techniques addressing sensitivity and intrusion issues”. In: *Ergonomics* 27.10 (1984), S. 1033–1050 (siehe S. 12).
- [CHN16] Stephen M. Casner, Edwin L. Hutchins und Don Norman. “The challenges of partially automated driving”. In: *Communications of the ACM* 59.5 (2016), S. 70–77 (siehe S. 8).
- [CT12] Michele Chinosi und Alberto Trombetta. “BPMN: An introduction to the standard”. In: *Computer Standards & Interfaces* 34.1 (2012), S. 124–134 (siehe S. 27).
- [CW02] Klaus Christoffersen und David D. Woods. “How to make automated systems team players”. In: *Advances in human performance and cognitive engineering research*. Hrsg. von D. Stone. Emerald Group Publishing Limited, 2002, S. 1–12 (siehe S. 8).
- [Cin+13] Burcu Cinaz, Bert Arnrich, Roberto La Marca und Gerhard Tröster. “Monitoring of mental workload levels during an everyday life office-work scenario”. In: *Personal and Ubiquitous Computing* 17.2 (2013), S. 229–239 (siehe S. 15).
- [Coh+18] Jeffrey F. Cohn, Nicholas Cummins, Julien Epps u. a. “Multimodal assessment of depression from behavioral signals”. In: *The Handbook of Multimodal-Multisensor Interfaces: Foundations, User Modeling, and Common Modality Combinations - Volume 2*. Hrsg. von Sharon Oviatt, Björn Schuller, Philip R. Cohen Cohen u. a. Association for Computing Machinery, 2018, S. 375–417 (siehe S. 20).
- [CL14] Jeffrey F. Cohn und Fernando de La Torre. “Automated Face Analysis for Affective”. In: *The Oxford Handbook of Affective Computing* (2014), S. 131 (siehe S. 18).
- [Coh04] Mike Cohn. *User stories applied: For agile software development*. Addison-Wesley Professional, 2004 (siehe S. 11).
- [CT92] T. F. Cootes und C. J. Taylor. “Active Shape Models — ‘Smart Snakes’”. In: *BMVC92*. Hrsg. von David Hogg und Roger Boyle. London: Springer London, 1992, S. 266–275 (siehe S. 43).
- [Cra+04] Scotty Craig, Arthur Graesser, Jeremiah Sullins und Barry Gholson. “Affect and learning: An exploratory look into the role of affect in learning with AutoTutor”. In: *Journal of Educational Media* 29.3 (2004), S. 241–250 (siehe S. 21).
- [CC06] D. Cristinacce und T. F. Cootes. “Feature Detection and Tracking with Constrained Local Models”. In: *Proceedings of the British Machine Vision Conference 2006*. Hrsg. von Mike Chantler, Bob Fisher und Manuel Trucco. British Machine Vision Association, 2006, S. 95.1–95.10 (siehe S. 43).

- [CD07] Padraig Cunningham und Sarah Jane Delany. “k-Nearest neighbour classifiers”. In: *Multiple Classifier Systems* 34.8 (2007), S. 1–17 (siehe S. 82).
- [DPG07] Sidney D’Mello, Rosalind W. Picard und Arthur Graesser. “Toward an Affect-Sensitive AutoTutor”. In: *IEEE Intelligent Systems* 4 (2007), S. 53–61 (siehe S. 21).
- [DT05] N. Dalal und B. Triggs. “Histograms of Oriented Gradients for Human Detection”. In: *CVPR 2005*. Hrsg. von Cordelia Schmid, Carlo Tomasi und Stefano Soatto. Los Alamitos, Calif: IEEE Computer Society, 2005, S. 886–893 (siehe S. 42, 46).
- [Dem06] Janez Demšar. “Statistical comparisons of classifiers over multiple data sets”. In: *Journal of Machine Learning Research* 7.Jan (2006), S. 1–30 (siehe S. 86).
- [Deu+15] Jochen Deuse, Kirsten Weisner, André Hengstebeck und Felix Busch. “Gestaltung von Produktionssystemen im Kontext von Industrie 4.0”. In: *Zukunft der Arbeit in Industrie 4.0*. Hrsg. von Alfons Botthof und Ernst Andreas Hartmann. Berlin und Heidelberg: Springer Berlin Heidelberg und Springer Vieweg, 2015, S. 99–109 (siehe S. 7).
- [Don+11] Yanchao Dong, Zhencheng Hu, Keiichi Uchimura und Nobuki Murayama. “Driver Inattention Monitoring System for Intelligent Vehicles: A Review”. In: *IEEE Transactions on Intelligent Transportation Systems* 12.2 (2011), S. 596–614 (siehe S. 20).
- [Du+15] Xiaoping Du, Yijing Zhang, Bin Wu u. a. “How Automation Effect Mental Workload of Novice Operators in Space Rendezvous and Docking”. In: *Engineering psychology and cognitive ergonomics*. Hrsg. von Don Harris. Lecture notes in computer science Lecture notes in artificial intelligence. Cham: Springer International Publishing und Springer, 2015, S. 317–326 (siehe S. 15).
- [Dum+15] Roman Dumitrescu, Jürgen Gausemeier, Arno Kühn u. a. “Auf dem Weg zu Industrie 4.0: Erfolgsfaktor Referenzarchitektur”. it’s OWL Clustermanagement GmbH, 2015 (siehe S. 6).
- [Dur+14] G. Durantin, J.-F. Gagnon, S. Tremblay und F. Dehais. “Using near infrared spectroscopy and heart rate variability to detect mental overload”. In: *Behavioural Brain Research* 259 (2014), S. 16–23 (siehe S. 15).
- [EFA80] Paul Ekman, Wallace V. Friesen und Sonia Ancoli. “Facial signs of emotional experience”. In: *Journal of personality and social psychology* 39.6 (1980), S. 1125 (siehe S. 19).
- [Eyb+10] Florian Eyben, Martin Wöllmer, Tony Poitschke u. a. “Emotion on the Road—Necessity, Acceptance, and Feasibility of Affective Computing in the Car”. In: *Advances in Human-Computer Interaction* 2010.1768 (2010), S. 1–17 (siehe S. 20).

- [Faw06] Tom Fawcett. “An introduction to ROC analysis”. In: *Pattern Recognition Letters* 27.8 (2006), S. 861–874 (siehe S. 83–85).
- [Fin93] H. Finner. “On a Monotonicity Problem in Step-Down Multiple Test Procedures”. In: *Journal of the American Statistical Association* 88.423 (1993), S. 920–923 (siehe S. 86, 90).
- [Fin10] Kraig Finstad. “The usability metric for user experience”. In: *Interacting with Computers* 22.5 (2010), S. 323–327 (siehe S. 13).
- [FM05] Elgar Fleisch und Friedemann Mattern. *Das Internet der Dinge*. Berlin/Heidelberg: Springer-Verlag, 2005 (siehe S. 6).
- [Fri91] Alan J. Fridlund. “Sociality of solitary smiling: Potentiation by an implicit audience”. In: *Journal of personality and social psychology* 60.2 (1991), S. 229 (siehe S. 18).
- [Fri01] Jerome H. Friedman. “Greedy Function Approximation: A Gradient Boosting Machine”. In: *The Annals of Statistics* 29.5 (2001), S. 1189–1232 (siehe S. 82).
- [FE78] E. Friesen und P. Ekman. “Facial action coding system: A technique for the measurement of facial movement”. In: *Palo Alto* (1978) (siehe S. 18).
- [Gar+10] Salvador García, Alberto Fernández, Julián Luengo und Francisco Herrera. “Advanced nonparametric tests for multiple comparisons in the design of experiments in computational intelligence and data mining: Experimental analysis of power”. In: *Information Sciences* 180.10 (2010), S. 2044–2064 (siehe S. 86, 90).
- [GEW06] Pierre Geurts, Damien Ernst und Louis Wehenkel. “Extremely randomized trees”. In: *Machine Learning* 63.1 (2006), S. 3–42 (siehe S. 82).
- [GS01] Richard Grace und Sonya Steward. “Drowsy driver monitor and warning system”. In: *International driving symposium on human factors in driver assessment, training and vehicle design*. Bd. 8. 2001, S. 201–208 (siehe S. 20).
- [Gra+13] Joseph F. Grafsgaard, Joseph B. Wiggins, Kristy Elizabeth Boyer, Eric N. Wiebe und James C. Lester. “Automatically Recognizing Facial Indicators of Frustration: A Learning-centric Analysis”. In: *2013 Humaine Association Conference on Affective Computing and Intelligent Interaction*. IEEE, 2013, S. 159–165 (siehe S. 21).
- [GK13] Joseph Greve und Annette Kluge. *Die Messung von Mental Workload und Situation Awareness während des Erlernens einer komplexen Prozesskontrolltätigkeit im Rahmen einer Eyetracker-Untersuchung*. 2013 (siehe S. 14).

- [Gri+07] Michael Grimm, Kristian Kroschel, Helen Harris u. a. “On the Necessity and Feasibility of Detecting a Driver’s Emotional State While Driving”. In: *Affective computing and intelligent interaction*. Hrsg. von Ana Paiva, Rui Prada und Rosalind W. Picard. Bd. 4738. Lecture Notes in Computer Science. Berlin: Springer, 2007, S. 126–138 (siehe S. 20).
- [HI66] Ernest A. Haggard und Kenneth S. Isaacs. “Micromomentary facial expressions as indicators of ego mechanisms in psychotherapy”. In: *Methods of Research in Psychotherapy*. Hrsg. von Louis A. Gottschalk und Arthur H. Auerbach. The Century Psychology Series. Boston, MA: Springer US, 1966, S. 154–165 (siehe S. 18).
- [HC18] Zakia Hammal und Jeffrey F. Cohn. “Automatic, Objective, and Efficient Measurement of Pain Using Automated Face Analysis”. In: *Social and Interpersonal Dynamics in Pain: We Don’t Suffer Alone*. Hrsg. von Tine Vervoort, Kai Karos, Zina Trost und Kenneth M. Prkachin. Cham: Springer International Publishing, 2018, S. 121–146 (siehe S. 20).
- [Har06] Sandra G. Hart. “Nasa-Task Load Index (NASA-TLX); 20 Years Later”. In: *Proceedings of the Human Factors and Ergonomics Society Annual Meeting 50.9* (2006), S. 904–908 (siehe S. 12).
- [HS88] Sandra G. Hart und Lowell E. Staveland. “Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research”. In: *Advances in psychology*. Hrsg. von Peter A. Hancock and Najmedin Meshkati. Bd. Volume 52. Amsterdam [u.a.]: Elsevier, 1988, S. 139–183 (siehe S. 12).
- [Har+18] Christian Harteis, Christoph Fischer, Torben Töniges und Britta Wrede. “Do we betray our errors beforehand? The use of eye tracking, automated face recognition and computer algorithms to analyse learning from errors”. In: *Frontline Learning Research* (2018) (siehe S. 121, 124).
- [HT06] Marc Hassenzahl und Noam Tractinsky. “User experience - a research agenda”. In: *Behaviour & Information Technology* 25.2 (2006), S. 91–97 (siehe S. 11, 17).
- [HP05] J. A. Healey und R. W. Picard. “Detecting Stress During Real-World Driving Tasks Using Physiological Sensors”. In: *IEEE Transactions on Intelligent Transportation Systems* 6.2 (2005), S. 156–166 (siehe S. 20).
- [Hea+98] M. A. Hearst, S. T. Dumais, E. Osuna, J. Platt und B. Scholkopf. “Support vector machines”. In: *IEEE Intelligent Systems and their Applications* 13.4 (1998), S. 18–28 (siehe S. 82).
- [HR11] Joel A. Hesch und Stergios I. Roumeliotis. “A Direct Least-Squares (DLS) method for PnP”. In: *IEEE International Conference on Computer Vision (ICCV), 2011*. Piscataway, NJ: IEEE, 2011, S. 383–390 (siehe S. 45).

- [HBK95] Ursula Hess, Rainer Banse und Arvid Kappas. “The intensity of facial expression is determined by underlying affective state and social situation”. In: *Journal of personality and social psychology* 69.2 (1995), S. 280 (siehe S. 18).
- [Hir15a] Hartmut Hirsch-Kreinsen. “Einleitung: Digitalisierung industrieller Arbeit”. In: *Digitalisierung industrieller Arbeit*. Hrsg. von Hartmut Hirsch-Kreinsen, Peter Ittermann und Jonathan Niehaus. Edition Sigma. Baden-Baden: Nomos, 2015, S. 9–30 (siehe S. 5, 9).
- [Hir15b] Hartmut Hirsch-Kreinsen. “Entwicklungsperspektiven von Produktionsarbeit”. In: *Zukunft der Arbeit in Industrie 4.0*. Hrsg. von Alfons Botthof und Ernst Andreas Hartmann. Berlin und Heidelberg: Springer Berlin Heidelberg und Springer Vieweg, 2015, S. 89–98 (siehe S. 9).
- [Hir14] Hartmut Hirsch-Kreinsen. *Wandel von Produktionsarbeit - "Industrie 4.0"*. TU, 2014 (siehe S. 7).
- [HL62] J. L. Hodges und E. L. Lehmann. “Rank Methods for Combination of Independent Experiments in Analysis of Variance”. In: *Ann. Math. Statist.* 33.2 (1962), S. 482–497 (siehe S. 86, 90).
- [Hör+16] Timm Hörmann, Marc Hesse, Peter Christ u. a. “Fine-Grained Prediction of Cognitive Workload in a Modern Working Environment by Utilizing Short-Term Physiological Parameters”. In: *Proceedings of the 9th International Joint Conference on Biomedical Engineering Systems and Technologies*. SCITEPRESS - Science and Technology Publications, 2016, S. 42–51 (siehe S. 15).
- [Hud03] Eva Hudlicka. “To feel or not to feel: The role of affect in human–computer interaction”. In: *Applications of Affective Computing in Human-Computer Interaction* 59.1–2 (2003), S. 1–32 (siehe S. 18).
- [Int10] International Organization for Standardization. *ISO 9241-210 Ergonomics of human-system interaction: Part 210: Human-centred design for interactive systems*, 2010 (siehe S. 10).
- [its18] it’s OWL Clustermanagement GmbH. *Digitale Transformation – Wir machen den Mittelstand fit! Das Technologie-Netzwerk: Intelligente Technische Systeme OstWestfalenLippe*. 2018 (siehe S. 25).
- [IN15] Peter Ittermann und Jonathan Niehaus. “Industrie 4.0 und Wandel von Industriearbeit: Überblick über Forschungsstand und Trendbestimmungen”. In: *Digitalisierung industrieller Arbeit*. Hrsg. von Hartmut Hirsch-Kreinsen, Peter Ittermann und Jonathan Niehaus. Edition Sigma. Baden-Baden: Nomos, 2015, S. 33–51 (siehe S. 7–9).

- [Jab+11] M. E. Jabon, J. N. Bailenson, E. Pontikakis, L. Takayama und C. Nass. "Facial-expression analysis for predicting unsafe driving behavior". In: *IEEE Pervasive Computing* 10.4 (2011), S. 84–95 (siehe S. 20).
- [Jai+15] Ashesh Jain, Hema S. Koppula, Bharad Raghavan, Shane Soh und Ashutosh Saxena. "Car That Knows Before You Do: Anticipating Maneuvers via Learning Temporal Driving Models". In: *ICCV 2015*. Hrsg. von Lisa O’Conner. Los Alamitos, CA [u.a.]: IEEE Computer Society, 2015, S. 3182–3190 (siehe S. 20).
- [JMF99] Esther Jakobs, Antony S. R. Manstead und Agneta H. Fischer. "Social Motives and Emotional Feelings as Determinants of Facial Displays: The Case of Smiling". In: *Personality and Social Psychology Bulletin* 25.4 (1999), S. 424–435 (siehe S. 18).
- [Joh+11] Hideo Joho, Jacopo Staiano, Nicu Sebe und Joemon M. Jose. "Looking at the viewer: Analysing facial activity to detect personal highlights of multimedia contents". In: *Multimedia Tools and Applications* 51.2 (2011), S. 505–523 (siehe S. 22).
- [JSW02] Karin Joiko, Martin Schmauder und Gertrud Wolff. *Psychische Belastung und Beanspruchung im Berufsleben: Erkennen-gestalten*. Bundesanst. für Arbeitsschutz und Arbeitsmedizin, 2002 (siehe S. 12).
- [Kag17] Henning Kagermann. "Chancen von Industrie 4.0 nutzen". In: *Handbuch Industrie 4.0 Bd. 4*. Hrsg. von Birgit Vogel-Heuser, Thomas Bauernhansl und Michael Ten Hompel. Springer, 2017, S. 237–248 (siehe S. 9).
- [KR05] Rana El Kaliouby und Peter Robinson. "Real-Time Inference of Complex Mental States from Facial Expressions and Head Gestures". In: *Real-Time Vision for Human-Computer Interaction* (2005), S. 181–200 (siehe S. 19).
- [Kin15] Davis E. King. "Max-Margin Object Detection". In: *CoRR* abs/1502.00046 (2015) (siehe S. 42).
- [Kle+04] G. Klein, D. D. Woods, J. M. Bradshaw, R. R. Hoffman und P. J. Feltovich. "Ten Challenges for Making Automation a "Team Player" in Joint Human-Agent Activity". In: *IEEE Intelligent Systems* 19.06 (2004), S. 91–95 (siehe S. 8).
- [Kor14] Oliver Korn. *Context-aware assistive systems for augmented work: A framework using gamification and projection*. Institut für Visualisierung und Interaktive Systeme (VIS) der Universität Stuttgart, 2014 (siehe S. 13).

- [KN16] Teoh Kung-Keat und Jasmine Ng. “Confused, Bored, Excited? An Emotion Based Approach to the Design of Online Learning Systems”. In: *7th International Conference on University Learning and Teaching (InCULT 2014) Proceedings*. Hrsg. von Chan Yuen Fook, Gurnam Kaur Sidhu, Suthagar Narasuman, Lee Lai Fong und Shireena Basree Abdul Rahman. Singapore: Springer Singapore, 2016, S. 221–233 (siehe S. 21).
- [LSG12] Bruno Laeng, Sylvain Sirois und Gustaf Gredebäck. “Pupillometry: A window to the preconscious?” In: *Perspectives on psychological science* 7.1 (2012), S. 18–27 (siehe S. 15).
- [Lan+12] Christian Lang, Sven Wachsmuth, Marc Hanheide und Heiko Wersing. “Facial Communicative Signals”. In: *International Journal of Social Robotics* 4.3 (2012), S. 249–262 (siehe S. 18).
- [LHS08] Bettina Laugwitz, Theo Held und Martin Schrepp. “Construction and Evaluation of a User Experience Questionnaire”. In: *HCI and Usability for Education and Work*. Hrsg. von Andreas Holzinger. Berlin, Heidelberg: Springer Berlin Heidelberg, 2008, S. 63–76 (siehe S. 12).
- [LC12] Boon-Giin Lee und Wan-Young Chung. “Driver Alertness Monitoring Using Fusion of Facial Features and Bio-Signals”. In: *IEEE Sensors Journal* 12.7 (2012), S. 2416–2422 (siehe S. 20).
- [LRU14] Jure Leskovec, Anand Rajaraman und Jeffrey D. Ullman. *Mining of massive datasets*. 2. ed. Cambridge [u.a.]: Cambridge Univ. Press, 2014 (siehe S. 6).
- [LRL07] Yulan Liang, Michelle L. Reyes und John D. Lee. “Real-Time Detection of Driver Cognitive Distraction Using Support Vector Machines”. In: *IEEE Transactions on Intelligent Transportation Systems* 8.2 (2007), S. 340–350 (siehe S. 20).
- [Luc+11] Patrick Lucey, Jeffrey F. Cohn, Iain Matthews u. a. “Automatically detecting pain in video through facial action units”. In: *IEEE transactions on systems, man, and cybernetics. Part B, Cybernetics : a publication of the IEEE Systems, Man, and Cybernetics Society* 41.3 (2011), S. 664–674 (siehe S. 20).
- [Lüd15] Andreas Lüdtkke. “Wege aus der Ironie in Richtung ernsthafter Automatisierung”. In: *Zukunft der Arbeit in Industrie 4.0*. Hrsg. von Alfons Botthof und Ernst Andreas Hartmann. Berlin und Heidelberg: Springer Berlin Heidelberg und Springer Vieweg, 2015, S. 125–146 (siehe S. 7, 8).
- [Lux16] Emma Luxton. *This factory is replacing robots with humans*. Hrsg. von World Economic Forum. <https://www.weforum.org/agenda/2016/02/this-factory-is-replacing-robots-with-humans/>, 2016 (siehe S. 6).
- [Mah36] Prasanta Chandra Mahalanobis. “On the generalized distance in statistics”. In: *Proceedings of the National Institute of Sciences of India*. 1936 (siehe S. 107).

- [Mal+11] Lucas Malta, Chiyomi Miyajima, Norihide Kitaoka und Kazuya Takeda. “Analysis of Real-World Driver’s Frustration”. In: *IEEE Transactions on Intelligent Transportation Systems* 12.1 (2011), S. 109–118 (siehe S. 20).
- [Mar+17] Brais Martinez, Michel F. Valstar, Bihan Jiang und Maja Pantic. “Automatic Analysis of Facial Actions: A Survey”. In: *IEEE Transactions on Affective Computing* (2017) (siehe S. 18).
- [McD+15] D. McDuff, R. E. Kaliouby, J. F. Cohn und R. W. Picard. “Predicting Ad Liking and Purchase Intent: Large-Scale Analysis of Facial Responses to Ads”. In: *IEEE Transactions on Affective Computing* 6.3 (2015), S. 223–235 (siehe S. 22).
- [MJ17] Natasha Merat und A. Hamish Jamson. “How Do Drivers Behave in a Highly Automated Car?” In: *Proceedings of the 5th International Driving Symposium on Human Factors in Driver Assessment, Training, and Vehicle Design : Driving Assessment 2009*. Iowa City, Iowa: University of Iowa, 2017, S. 514–521 (siehe S. 8).
- [MH88] N. Meshkati und P. A. Hancock. *Human Mental Workload*. 1. Aufl. Bd. v. 52. Advances in Psychology. s.l.: Elsevier textbooks, 1988 (siehe S. 12).
- [MF07] Daniel Messinger und Alan Fogel. “The interactive development of social smiling”. In: *Advances in child development and behaviour* 35 (2007), S. 328–366 (siehe S. 18).
- [Mij+16] Pavle Mijović, Vanja Ković, Maarten de Vos u. a. “Benefits of Instructed Responding in Manual Assembly Tasks: An ERP Approach”. In: *Frontiers in Human Neuroscience* 10 (2016), S. 171 (siehe S. 15, 16).
- [Nav+17] R. Navarathna, P. Carr, P. Lucey und I. Matthews. “Estimating Audience Engagement to Predict Movie Ratings”. In: *IEEE Transactions on Affective Computing* (2017), S. 1 (siehe S. 22).
- [Nor05] Don Norman. “Human-centered design considered harmful”. In: *interactions* 12.4 (2005), S. 14–19 (siehe S. 16).
- [Nor13] Don Norman. *The design of everyday things: Revised and expanded edition*. Constellation, 2013 (siehe S. 10).
- [ND86] Don Norman und Stephen Draper. *User Centered System Design; New Perspectives on Human-Computer Interaction*. Hillsdale, NJ, USA: L. Erlbaum Associates Inc, 1986 (siehe S. 10).
- [NMH95] Don Norman, Jim Miller und Austin Henderson. “What you see, some of what’s in the future, and how we go about doing it HI at Apple Computer”. In: *Conference companion on Human factors in computing systems*. Hrsg. von ACM. New York, NY, USA: ACM, 1995, S. 155 (siehe S. 11).

- [NBS17] David Nunes, Fernando Boavida und Jorge Sá Silva. *A practical introduction to human-in-the-loop cyber-physical systems*. First edition. Hoboken, NJ: John Wiley & Sons, 2017 (siehe S. 7).
- [Obj09] Object Management Group. *Business Process Model and Notation (BPMN) Specification Version 2.0*. 2009 (siehe S. 27).
- [Ols+16] Randal S. Olson, Ryan J. Urbanowicz, Peter C. Andrews u. a. “Automating Biomedical Data Science Through Tree-Based Pipeline Optimization: Applications of Evolutionary Computation: 19th European Conference, EvoApplications 2016, Porto, Portugal, March 30 – April 1, 2016, Proceedings, Part I”. In: *Applications of Evolutionary Computation: 19th European Conference, EvoApplications 2016, Porto, Portugal, March 30 – April 1, 2016, Proceedings, Part I*. Hrsg. von Giovanni Squillero und Paolo Burelli. Springer International Publishing, 2016, S. 123–137 (siehe S. 78, 79).
- [PGB14] Julie Paxion, Edith Galy und Catherine Berthelon. “Mental workload and driving”. In: *Frontiers in psychology* 5 (2014), S. 1344 (siehe S. 20).
- [Ped+11] F. Pedregosa, G. Varoquaux, A. Gramfort u. a. “Scikit-learn: Machine Learning in Python”. In: *Journal of Machine Learning Research* 12 (2011), S. 2825–2830 (siehe S. 79).
- [PBX09] Jian Peng, Liefeng Bo und Jinbo Xu. “Conditional Neural Fields”. In: *Advances in Neural Information Processing Systems 22*. Hrsg. von Y. Bengio, D. Schuurmans, J. D. Lafferty, C. K. I. Williams und A. Culotta. Curran Associates, Inc, 2009, S. 1419–1427 (siehe S. 44).
- [PVH01] R. W. Picard, E. Vyzas und J. Healey. “Toward machine emotional intelligence: analysis of affective physiological state”. In: *IEEE transactions on pattern analysis and machine intelligence* 23.10 (2001), S. 1175–1191 (siehe S. 18).
- [Pow11] David Martin Powers. *Evaluation: From precision, recall and F-measure to ROC, informedness, markedness and correlation*. Bioinfo Publications, 2011 (siehe S. 83).
- [Pro13] Promotorengruppe Kommunikation der Forschungsunion Wirtschaft - Wissenschaft. *Umsetzungsempfehlungen für das Zukunftsprojekt Industrie 4.0*. 2013 (siehe S. 5, 6).
- [Qio+06] Qiong Wang, Jingyu Yang, Mingwu Ren und Yujie Zheng. “Driver Fatigue Detection: A Survey”. In: *WCICA 2006*. IEEE, 2006, S. 8587–8591 (siehe S. 20).

- [Ram+12] Beatrice Rammstedt, Christoph J. Kemper, Mira Céline Klein, Constanze Beierlein und Anastassiya Kovaleva. “Eine kurze Skala zur Messung der fünf Dimensionen der Persönlichkeit: Big-Five-Inventory-10 (BFI-10)”. In: *GESIS-Working Papers, 2012/23, GESIS - Leibniz-Institut für Sozialwissenschaften* (2012) (siehe S. 73, 143, 159).
- [Rei03] Frederick F. Reichheld. “The one number you need to grow”. In: *Harvard business review* 81.12 (2003), S. 46–55 (siehe S. 13).
- [RN88] Gary B. Reid und Thomas E. Nygren. “The subjective workload assessment technique: A scaling procedure for measuring mental workload”. In: *Advances in psychology*. Hrsg. von Peter A. Hancock und Najmedin Meshkati. Bd. 52. Elsevier, 1988, S. 185–218 (siehe S. 12).
- [Röb17] Peter Röben. “Industrie 4.0 und Berufsbildung 4.0: Zwei Projekte, eine Nummer?” In: *Einheit und Differenz in den gewerblich-technischen Wissenschaften*. Hrsg. von Matthias Becker, Christian Dittmann, Julia Gillen, Stefanie Hiestand und Rita Meyer. Bd. 33. Bildung und Arbeitswelt. Berlin und Münster: LIT Verlag, 2017, S. 246–267 (siehe S. 5, 6).
- [Rod+15] I. Rodríguez-Fdez, A. Canosa, M. Mucientes und A. Bugarín. “STAC: A web platform for the comparison of algorithms using statistical tests”. In: *2015 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*. 2015, S. 1–8 (siehe S. 86).
- [Rom+16] David Romero, Peter Bernus, Ovidiu Noran, Johan Stahre und Åsa Fast-Berglund. “The Operator 4.0: Human Cyber-Physical Systems & Adaptive Automation Towards Human-Automation Symbiosis Work Systems”. In: *Advances in Production Management Systems. Initiatives for a Sustainable World*. Hrsg. von Irenilza Nääs, Oduvaldo Vendrametto, João Mendes Reis u. a. Cham: Springer International Publishing, 2016, S. 677–686 (siehe S. 7, 8).
- [Rou84] Peter J. Rousseeuw. “Least Median of Squares Regression”. In: *Journal of the American Statistical Association* 79.388 (1984), S. 871–880 (siehe S. 106).
- [Rv99] Peter J. Rousseeuw und Katrien van Driessen. “A Fast Algorithm for the Minimum Covariance Determinant Estimator”. In: *Technometrics* 41.3 (1999), S. 212–223 (siehe S. 106).
- [Rüt+13] Stefan Rütter, Thomas Hermann, Maik Mracek, Stefan Kopp und Jochen Steil. “An Assistance System for Guiding Workers in Central Sterilization Supply Departments”. In: *Proceedings of the 6th International Conference on Pervasive Technologies Related to Assistive Environments*. Hrsg. von Fillia Makedon, Margrit Betke, Magy Self El-Nasr und Ilias Maglogiannis. PETRA '13. New York, NY, USA: ACM, 2013, 3:1–3:8 (siehe S. 27).

- [Ryz18] Paul Ryznar. *Bye, Robot? Bringing Back Human Workers Bucks Manufacturing Trends*. Hrsg. von Forbes. <https://www.forbes.com/sites/forbestechcouncil/2018/03/26/bye-robot-bringing-back-human-workers-bucks-manufacturing-trends/>, 2018 (siehe S. 6).
- [SLC11] Jason M. Saragih, Simon Lucey und Jeffrey F. Cohn. “Deformable Model Fitting by Regularized Landmark Mean-Shift”. In: *International Journal of Computer Vision* 91.2 (2011), S. 200–215 (siehe S. 44).
- [Sow+16] Sulayman K. Sowe, Koji Zettsu, Eric Simmon, Frederic de Vault und Irena Bojanova. “Cyber-Physical Human Systems: Putting People in the Loop”. In: *IT professional* 18.1 (2016), S. 10–13 (siehe S. 7).
- [SC08] Ingo Steinwart und Andreas Christmann. *Support vector machines*. Springer Science & Business Media, 2008 (siehe S. 82).
- [SKH17] Adam Szulewski, Danielle Kelton und Daniel Howes. “Pupillometry as a Tool to Study Expertise in Medicine”. In: *Frontline Learning Research* 5.3 (2017), S. 53–63 (siehe S. 15, 123).
- [TT05] Jianhua Tao und Tieniu Tan. “Affective Computing: A Review”. In: *Affective computing and intelligent interaction*. Hrsg. von Jianhua Tao, Tieniu Tan und Rosalind W. Picard. Lecture Notes in Computer Science. Berlin: Springer Berlin Heidelberg und Springer, 2005, S. 981–995 (siehe S. 18).
- [Tho+17] Jean-Paul Thommen, Ann-Kristin Achleitner, Dirk Ulrich Gilbert, Dirk Hachmeister und Gernot Kaiser. *Allgemeine Betriebswirtschaftslehre*. 8. Auflage. Lehrbuch. Wiesbaden: Springer Gabler, 2017 (siehe S. 22).
- [Tön+16] Torben Töniges, Sonja Ötting, Britta Wrede, Günter W. Maier und Gerhard Sagerer. “An emerging decision authority: Adaptive cyber-physical system design for fair human-machine interaction and decision processes”. In: *Cyber-physical systems: Foundations, principles, and applications* (2016) (siehe S. 9).
- [UHM11] Dieter Uckelmann, Mark Harrison und Florian Michahelles. *Architecting the Internet of Things*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011 (siehe S. 6).
- [VPB09] Alessandro Vinciarelli, Maja Pantic und Hervé Bourlard. “Social signal processing: Survey of an emerging domain”. In: *Visual and multimodal analysis of human spontaneous behaviour* 27.12 (2009), S. 1743–1759 (siehe S. 18).
- [Vre+02] Karel Vredenburg, Ji-Ye Mao, Paul W. Smith und Tom Carey. “A Survey of User-centered Design Practice”. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. Hrsg. von ACM. CHI '02. New York, NY, USA: ACM, 2002, S. 471–478 (siehe S. 10).

- [Vur+07] Esra Vural, Mujdat Cetin, Aytul Ercil u. a. “Drowsy Driver Detection Through Facial Movement Analysis”. In: *Human-computer interaction*. Hrsg. von Michael Lew, Nicu Sebe, Thomas S. Huang und Erwin M. Bakker. Bd. 4796. Lecture Notes in Computer Science. Berlin: Springer, 2007, S. 6–18 (siehe S. 20).
- [WBJ67] Paul Watzlawick, Janet B. Bevelas und Don D. Jackson. *Pragmatics of Human Communication*. WW Norton & Company, 1967 (siehe S. 2, 17).
- [Whi+14] Jacob Whitehill, Zewelangi Serpell, Yi-Ching Lin, Aysha Foster und Javier R. Movellan. “The Faces of Engagement: Automatic Recognition of Student Engagement from Facial Expressions”. In: *IEEE Transactions on Affective Computing* 5.1 (2014), S. 86–98 (siehe S. 21).
- [Wil45] Frank Wilcoxon. “Individual Comparisons by Ranking Methods”. In: *Biometrics Bulletin* 1.6 (1945), S. 80–83 (siehe S. 86).
- [Win14] Lars Windelband. “Zukunft der Facharbeit im Zeitalter „Industrie 4.0 “”. In: *Journal of Technical Education (JOTED)* 2.2 (2014) (siehe S. 7).
- [Wis14] Wissenschaftliche Beirat der Plattform Industrie 4.0. “Neue Chancen für unsere Produktion: 17 Thesen des wissenschaftlichen Beitrats der Plattform Industrie 4.0”. 2014 (siehe S. 7).
- [Woo+15] Erroll Wood, Tadas Baltrusaitis, Xucong Zhang u. a. “Rendering of Eyes for Eye-Shape Registration and Gaze Estimation”. In: *The IEEE International Conference on Computer Vision (ICCV)*. 2015 (siehe S. 45).
- [Yuc+17] Anil Yuce, Hua Gao, Gabriel L. Cuendet und Jean-Philippe Thiran. “Action Units and Their Cross-Correlations for Prediction of Cognitive Load during Driving”. In: *IEEE Transactions on Affective Computing* 8.2 (2017), S. 161–175 (siehe S. 20).
- [Zen+09] Zhihong Zeng, Maja Pantic, Glenn I. Roisman und Thomas S. Huang. “A Survey of Affect Recognition Methods: Audio, Visual, and Spontaneous Expressions”. In: *IEEE transactions on pattern analysis and machine intelligence* 31.1 (2009), S. 39–58 (siehe S. 2).

Kolophon

Diese Promotion wurde mit L^AT_EX 2_ε gesetzt. Dabei wurde die *Clean Thesis* Stilvorlage von Ricardo Langner verwendet die von Nutzeranleitungsdokumenten von Apple Inc. inspiriert wurde.

Die Stilvorlage kann auf <http://cleanthesis.der-ric.de/> heruntergeladen werden.