TECHNISCHE UNIVERSITÄT
CHEMNITZ

Fakultät für Informatik

CSR-18-03

StayCentered - Methodenbasis eines Assistenzsystems für Centerlotsen (MACeLot) Schlussbericht

Guido Brunnett · Maximilian Eibl · Fred Hamker ·
Peter Ohler · Peter Protzel

November 2018

Chemnitzer Informatik-Berichte

StayCentered – Methodenbasis eines Assistenzsystems für Centerlotsen

(MACeLot)

Schlussbericht

Verbundprojekt im Rahmen des Innovationsfeldes „Mensch-Technik-Interaktion für den demographischen Wandel“ im Rahmen des Förderprogramms „IKT – Forschung für Innovation“ des Bundesministeriums für Bildung und Forschung (BMBF)

Autorinnen und Autoren: **Prof. Dr. Guido Brunnett, Prof. Dr. Maximilian Eibl**

Prof. Dr. Fred Hamker, Prof. Dr. Peter Ohler, Prof. Dr.-Ing. Peter Protzel

Technische Universität Chemnitz

Straße der Nationen 62

09107 Chemnitz

TECHNISCHE UNIVERSITÄT CHEMNITZ

Interdisziplinäres Kompetenzzentrum **Virtual Humans**

VDI|VDE|IT

Bundesministerium für Bildung und Forschung

Zuwendungsempfänger

Technische Universität Chemnitz
Interdisziplinäre Kompetenzzentrum ‚Virtual Humans'

Projektträger
VDI/VDE-IT GmbH
Förderkennzeichen:
16SV7260

Vorhabenbezeichnung
StayCentered – Methodenbasis eines Assistenzsystems für Centerlotsen

Gesamtlaufzeit:
01. Februar 2015 – 30. April 2018

Berichtszeitraum:
01. Februar 2015 – 30. April 2018

Datum der Fälligkeit des Berichts:
14. November 2018

Kontakt
Prof. Dr. Guido Brunnett
Professur Graphische Datenverarbeitung und Visualisierung
Technische Universität Chemnitz
Straße der Nationen 62
09107 Chemnitz
E-Mail: guido.brunnett@informatik.tu-chemnitz.de
Web: https://www.tu-chemnitz.de/forschung/virtual_humans/

Haftungsausschluss

Inhalt

I. Kurzdarstellung

1. Aufgabenstellung

Die Qualität kooperativer Arbeit an technischen Systemen wird wesentlich durch die Qualität der Kommunikation und Interaktion zwischen den handelnden Personen bestimmt. Die Absicherung von Arbeitsqualität und -produktivität sowie die Vermeidung von Systemausfällen durch Fehlbedienung beinhalten somit sowohl technische als auch psychologische und soziologische Fragestellungen. Insbesondere in Tätigkeitsfeldern, bei denen menschliches Versagen weitreichende Konsequenzen haben kann, besteht deshalb ein erhebliches Interesse an der Entwicklung von Assistenzsystemen zur Unterstützung des kooperativen Arbeitsprozesses. Von einem solchen System wird erwartet, dass es den emotionalen Zustand der Individuen und den Gesamtzustand des Teams beurteilen und in Grenzen auch vorhersagen kann, um kritische Situationen im Vorfeld zu vermeiden. Darüber hinaus sollte das System den Benutzern Informationen bereitstellen, um den Arbeitsprozess effektiver zu gestalten, aber auch Metadaten generieren, die Schwachstellen des Arbeitsplatzentwurfes deutlich machen oder Hinweise darauf geben, welche Zusammenstellung der Arbeitsteams zu einer größeren Leistungsfähigkeit und Arbeitszufriedenheit führen kann.

Die Aufgabenstellung des Projektes bestand in der Entwicklung des notwendigen Methodenvorrates zur Realisierung eines emotionssensitiven Assistenzsystems, welches das Personal an einem kooperativen Arbeitsplatz situativ unterstützt. Darüber hinaus sollte das System für mögliche Supervisoren Informationen über den Zustand der Belastung der Teams liefern und Metadaten generieren, die Schwachstellen des Arbeitsplatzentwurfes deutlich machen. Die realisierten Konzepte und Verfahren wurden im Sinne eines Nachweises der prinzipiellen Funktionsfähigkeit (Proof-of-Concept) für den Arbeitsplatz von Centerlotsen prototypisch umgesetzt. Diese Umsetzung war notwendig, da das Zusammenspiel einzelner Verfahren (etwa bei der Sensorfusion) einen wesentlichen Einfluss auf die Funktionsfähigkeit des Gesamtsystems besitzt und daher anwendungsorientiert getestet werden sollte.

Abbildung 1 gibt einen Überblick über die Struktur des Gesamtsystems und seine Hauptbestandteile. Die zentrale Komponente des anvisierten Assistenzsystems leistet die Simulation des „internen Zustands“ des Teams in Bezug auf das Kommunikations- und Interaktionsverhalten und der zugrundeliegenden emotionalen Zustände der Teammitglieder. Hierzu wurde zunächst ein abstraktes Modell entwickelt, welches die möglichen Zustände der Emotionen und des Kommunikations- und Interaktionsverhaltens, die bestimmenden Größen dieser Zustände sowie die wechselseitigen Abhängigkeiten zwischen den Zuständen beschreibt. Das Simulationsmodul besteht aus einer Implementierung dieses Modells und einer Vorverarbeitung von Sensordaten, die aus den Rohdaten die vom Emotionsmodul benötigten gefilterten und synchronisierten Eingabedaten erzeugt.

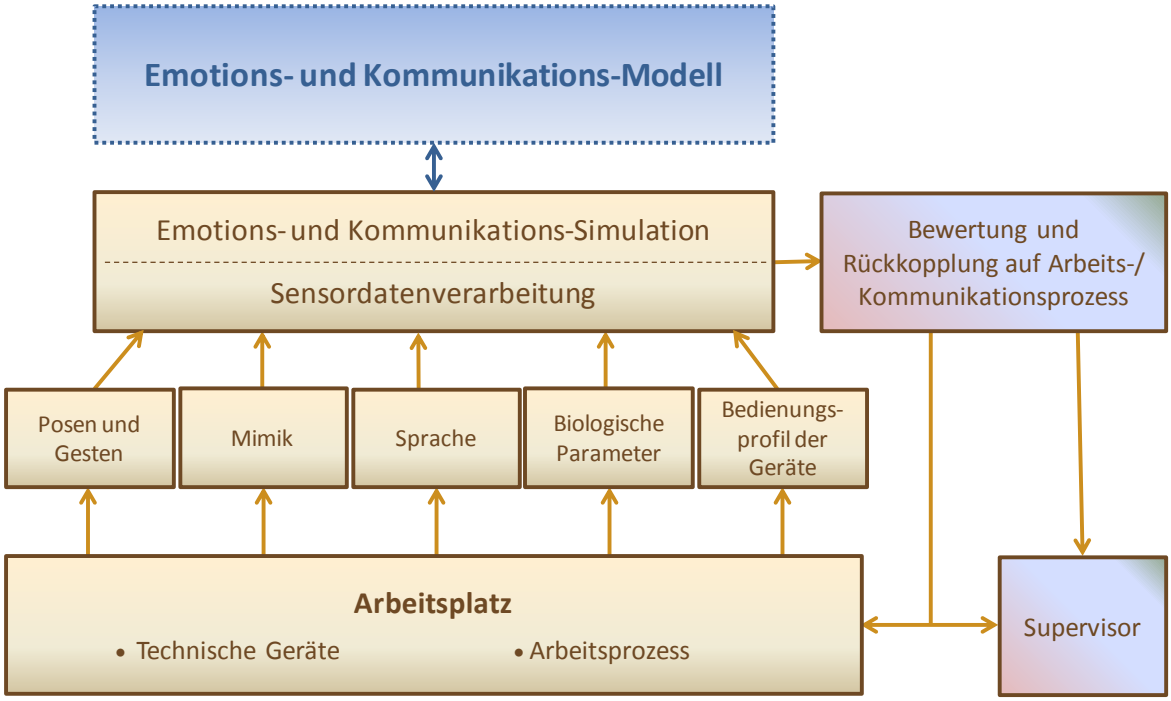


Abbildung 1: Struktur des Gesamtsystems

Das Simulationsmodul erkennt Tendenzen im Sozialverhalten der Akteure (z.B. Zunahme des Stressniveaus, Erhöhung des Aggressionspotentials) und errechnet durch Extrapolation dieser Tendenzen und konfigurierbarer Randbedingungen (z.B. gleichbleibendes Anforderungsniveau) eine mögliche Entwicklung des Kommunikationsverhaltens. Diese Vorhersage wird beständig korrigiert durch den eingehenden Strom von Sensordaten, die die Grundlage zur Erfassung des aktuellen inneren Zustands des Teams bilden. Um möglichst umfassende Informationen über den aktuellen Zustand des Teams zu erhalten, wurden fünf verschieden Kanäle zur Datenerfassung realisiert:

1. Motion Capturing erlaubt die Erfassung von Körperhaltungen und Gesten der Teammitglieder. Durch Rekonstruktion ihrer räumlichen Situierung und Proxemik wird ein dreidimensionaler Rahmen der Kommunikationssituation geschaffen, dem alle weiteren Sensordaten zugeordnet werden.
2. Über das Tracking von Mimik-Features können Bewegungen im Gesicht des Nutzers erkannt werden. Dabei sind spezifische Action Units mit bestimmten emotionalen Reaktionen verknüpft (emFACS).
3. Durch die Analyse geeigneter Stimmen-Parameter (z.B. Tonhöhe, Geschwindigkeit, Sprechpausen, Energieverteilung im Stimmenspektrum) lassen sich ebenfalls emotionale Befindlichkeiten bestimmen.
4. Biologische Parameter wie Herzfrequenz, Blutdruck und Hautleitwiderstand geben direkte Hinweise auf Stressbelastung und emotionale Erregung.
5. Die in Logfiles speicherbaren Benutzerprofile der verwendeten Geräte gaben ebenfalls Hinweise auf vorliegende Belastungssituationen.

Die Systemreaktion erfolgt in Abhängigkeit der beobachteten Kommunikation, kann aber auch anwendungsspezifisch variieren. Die Untersuchung, welche Reaktionsform des Systems den Arbeits- und Kommunikationsprozess des Personals am besten unterstützt, inklusive der Festlegung der Medien zur Übermittlung von Informationen an den Menschen, war ebenfalls Bestandteil des Projektes.

2. Voraussetzungen, unter denen das Vorhaben durchgeführt wurde

MACeLot war ein Teilvorhaben des BMBF-Förderschwerpunktes „Sozial- und emotionssensitive Systeme“ (InterEmotio). Das Projekt wurde von sechs Professuren der TU Chemnitz durchgeführt, die als Mitglieder des Interdisziplinären Kompetenzzentrums *Virtual Humans* strukturell und inhaltlich vernetzt sind. Die Kooperationspartner verfügten über alle erforderlichen Kompetenzen, die durch einschlägige Vorarbeiten belegt wurden.

Professur Graphische Datenverarbeitung und Visualisierung (GDV, Leitung: Prof. Dr. Guido Brunnett): Die Forschungsschwerpunkte der Professur liegen auf den Gebieten der Verarbeitung geometrischer Informationen (Geometric Processing) und der Virtuellen Realität. Die Professur besitzt umfangreiche Erfahrungen in der Verarbeitung menschlicher Bewegungsdaten. Dies betrifft sowohl die Segmentierung und Klassifizierung als auch die Synthetisierung von Bewegungsdaten. Es wurden verschiedene Beiträge zur markerlosen Bewegungserfassung geleistet [123]. Im Rahmen des Projektes The Smart Virtual Worker wurde ein Bewegungsgenerator für die Verrichtung von Tätigkeiten in der industriellen Fertigung erstellt, der die benötigten Bewegungen durch Modifikation aufgezeichneter Bewegungsdaten erzeugt. Im DFG-Projekt BR 1185/13-1 wurden die Bewegungen eines realen Karatekämpfers in Echtzeit klassifiziert, um eine Konterbewegung des virtuellen Gegners zu berechnen. Im Rahmen des Graduiertenkollegs „Crossworlds“ wurde ein gestenbasiertes Interaktionskonzept für mehrbenutzerfähige immersive Umgebungen entwickelt.

Professur Medienpsychologie (MP, Leitung: Prof. Dr. Peter Ohler): Die Professur betreibt mit einer kognitionspsychologischen Orientierung umfangreiche empirische Studien zu Medienwirkung und -nutzung und zur Mensch-Computer-Interaktion mit einem Schwerpunkt in den Bereichen Kognition und Emotion. Dabei existieren langjährige Erfahrungen im Bereich der kognitiven Modellierung, z.B. mittels mentaler Modelle und mit Agentensystemen. Zwei aktuelle Schwerpunkte bilden räumliche Kognitionen und die multimodale Emotionserkennung bei natürlichen und künstlichen Aktoren. Über einen langen Zeitraum wurden jeweils innovative Verfahren zur Messung von Informationsverarbeitungsprozessen entwickelt. Die Arbeitsgruppe nutzt seit vielen Jahren physiologische Messungen und Logfile-Recordings z.B. im Bereich der Nutzung von CAD-Systemen. Ferner wurde die Messung von Blickbewegung als Online Maß für kognitive Prozesse in unterschiedlichen Domänen angewandt. Die Professur ist mit zwei Schwerpunkten (Sensomotorik und Emotion) im Graduiertenkolleg „CrossWorlds“ beteiligt, zudem entwickelt die Arbeitsgruppe in enger Zusammenarbeit mit der Arbeitsgruppe PA das Emotionsmodul im Projekt The Smart Virtual Worker, das im Zuge von MACeLot weiterentwickelt wurde.

Professur Prozessautomatisierung (PA, Leitung: Prof. Dr.-Ing. Peter Protzel): Die Professur hat ihren Forschungsschwerpunkt auf dem Gebiet der autonomen Systeme. Autonome Systeme erfassen ihre Umwelt mit verschiedenen Sensoren und treffen aufgrund von Steuer- und Regelalgorithmen selbstständig Entscheidungen und führen Aktionen aus, ohne dass der Mensch eingreift. Dazu gehören autonome mobile Roboter, Fahrerassistenzsysteme oder „intelligente“ Fabriken. Ein wesentlicher Bestandteil der Arbeiten im Bereich der mobilen Robotik bildet die Sensordatenverarbeitung und Auswertung für Regelungs- und Navigationsaufgaben. Mit Sensordatenfusion mittels Faktorgraphen beschäftigte sich die Dissertation [77], deren Ergebnisse in abgewandelter Form für das vorliegende Projekt genutzt werden konnte. Des Weiteren spielt die Modellierung von Emotionen und deren Vorhersage im Projekt The Smart Virtual Worker eine Schlüsselrolle. Das hierbei entwickelte Modell kann für unterschiedliche Konstitutionstypen einen emotionale Valenz- und deren Erregungslevel berechnen

([155][97]) und soll schrittweise zu einem möglichst allgemeinen Emotionsmodell weiter entwickelt werden, welches auch für andere Arbeitsanwendungen (z.B. E-Learning) angewendet werden kann.

Professur Medieninformatik (MI, Leitung: Prof. Dr. Maximilian Eibl): Die Professur beschäftigt sich seit 2007 mit der Analyse von audiovisuellen Medien, vor allem unter dem Aspekt der Archivierung und des Retrieval. In Nachwuchsforschergruppen wie SachsMedia (www.sachsmedia.tv), validAX (www.validax.de) oder Chrooma+ (www.chroomaplus.eu) wurden Verfahren zur Bild- und Audioanalyse entwickelt. Ein Schwerpunkt lag dabei auf der Analyse von low-level-Features im Audiosignal sowie Verfahren der automatischen Transkription von Gesprochenem. Neben zahlreichen nationalen und internationalen Publikationen sind auch mehrere Promotionen zum Thema entstanden. Die Professur nimmt regelmäßig mit sehr großem Erfolg an der EU-organisierten Evaluationskampagne CLEF (Conference & Labs of the Evaluation Forum, www.clef-initiative.eu) teil. Praktisch entstanden sind dabei das Framework AMOPA - *Automated Moving Picture Annotator*, welches die automatische Analyse von audiovisuellen Medien ermöglicht, sowie eine Annotationsoberfläche, die eine leichte und schnelle manuelle Annotation, unterstützt durch die automatischen Verfahren von AMOPA, erlaubt.

Professur Künstliche Intelligenz (KI, Leitung: Prof. Dr. Fred Hamker): Die Professur befasst sich in der Forschung und Lehre mit der Entwicklung von autonomen, kognitiven Systemen. Besonderer Schwerpunkt sind dabei visuelle Systeme. Die Methoden umfassen die neuronale Modellierung nach dem Vorbild der Funktionsweise des Gehirns und Algorithmen des Maschinellen Lernens. Die Professur arbeitet derzeit an verschiedenen DFG Projekten und koordinierte ein FET FP7 EU-Projekt (SpaceCog). Im Graduiertenkolleg „CrossWorlds“ bearbeitet die Professur an der Entwicklung von emotionalen (virtuellen) Agenten im Kontext einer Face-Face Kommunikation zwischen Mensch und Agent. Hierzu wird ein Sehsystem des Agenten modelliert, welches in der Lage sein soll, die Gesichtsausdrücke des Menschen unterscheiden zu können und so dem Agenten Information über den emotionalen Zustand seines Gegenübers bereitstellt.

Juniorprofessur Visual Computing (VC, Leitung: Jun.-Prof. Dr. Paul Rosenthal, bis September 2016 am Projekt beteiligt): Die Juniorprofessur fokussierte sich in der Forschung auf die Entwicklung von Visualisierungssystemen. Dabei standen einerseits interaktive Systemen zur intuitiven Visualisierung von großen und abstrakten Datenmengen im Vordergrund. Einige der Projekte untersuchten die intuitive Visualisierung von Airline Operationsdaten verbunden mit kollaborativer Interaktion in Crewdienstplänen (Kooperation mit der Lufthansa Systems AG), die gestenbasierte und dinghafte (tangible) Interaktion im interdisziplinären Kontext (Graduiertenkolleg „CrossWorlds“), die intuitive Visualisierung von zeitveränderlichen Arbeitsergonomiedaten (ESF-Projekt „The Smart Virtual Worker“) und die visuell intuitive Aufbereitung von Daten aus optischer Kohärenztomografie für eine Erstdiagnose durch niedergelassene Augenärzte (Kooperation mit der Novartis Pharma GmbH). Ein zweiter Bereich beschäftigte sich mit der Visualisierung von massiven Datensätzen und deren interaktiver Verarbeitung auf Grafikhardware, insbesondere Punktwolkenrendering (Eurostars Projekt „enercloud“) und Visualisierung von astrophysikalischen SPH Daten (DFG-Projekt „SmoothViz“).

3. Planung und Ablauf des Vorhabens

Das Projekt wurde vom 01. Februar 2015 bis 30. April 2018 durchgeführt. Die einzelnen Projektpartner übernahmen folgende Teilaufgaben:

- **Professur Graphische Datenverarbeitung und Visualisierung (GDV)**: Die Aufgabe dieser Professur bestand in der Erfassung und Klassifizierung kommunikationsrelevanter Posen und Gesten der Teammitglieder.
- **Professur Künstliche Intelligenz (KI):** Die Aufgabe dieser Professur bestand in der Erfassung und Klassifizierung von Gesichtsausdrücken.
- **Professur Medieninformatik (MI):** Die Aufgabe dieser Professur bestand in der Extraktion und Verarbeitung emotionaler Features in Sprachsignalen.
- **Professur Prozessautomatisierung (PA)**: Neben der Entwicklung des Emotions- und Kommunikationsmodells war die Professur für die Durchführung der Sensorfusion verantwortlich.
- **Professur Medienpsychologie (MP):** Die Aufgaben der Professur lagen in der gemeinsamen Entwicklung des Emotions- und Kommunikationsmodells mit der Professur Prozessautomatisierung (PA), der Evaluation der erzeugten Systemreaktion auf die Teams sowie die Auswertung von Bedienprofilen der genutzten Eingabegeräte.
- **Juniorprofessur Visual Computing (VC)**: Die Aufgabe dieser Professur bestand in der Konzeption und Realisierung geeigneter Systemreaktionen, insbesondere Anpassung der Benutzeroberfläche für die kooperierenden Nutzer und Visualisierung der Systemausgaben für die Supervisoren. Ab dem 01.10.2016 wurden die Aufgaben und Verantwortlichkeiten von der Professur Medienpsychologie übernommen, da die Juniorprofessur turnusgemäß zum 30.09.2016 auslief.
- **DFS Deutsche Flugsicherung GmbH (DFS):** Die DFS fungierte als Industriepartner und stellte Möglichkeiten bereit, Fluglotsen bei ihrer Arbeit zu untersuchen.

Die Bearbeitung der Aufgaben im Projekt erfolgte gemäß dem folgenden Ablaufplan. Die Ergebnisse des Projektes sind in Abschnitt II.1 dargestellt.

Schwer-punkt	Lfd. Nr.	Arbeitspaket (AP)	Hauptverant-wortlich	Weitere Beteiligte	PM	Bearbeitungs-zeitraum
Modellierung	2.1	Datengenerierung und Aufbereitung für Sensordatenfusion	PA	alle	3	M 1-3
	2.2	Konzeption Sensordatenfusion	PA	alle	3	M 4-6
	2.3	Implementierung der Sensordatenfusion	PA	keine	6	M 7-12
	16.1	Einbindung aller sensorischen Eingabemodule	PA	alle	4	M 25-28
	16.2	Optimierung Sensordatenfusion an Reaktionen Input	PA	keine	4	M 29-32
	16.3	Evaluation der Sensordatenfusion mit kognitiven Kommunikationsmodell	PA	keine	4	M 33-36
	8.1	Entwicklung des kognitiven Kommunikations-modells	MP+PA	keine	4	M 13-16
	8.2	Prototypen-Datenintegration	MP+PA	keine	2	M 17-18
	8.3	Modell-Realitätsabgleich	MP+PA	keine	3	M 19-21
	8.4	Prototypen-Implementierung	MP+PA	keine	3	M 22-24
	6.1	Aufarbeitung wissenschaftlicher Stand und Design der empirischen Erhebu	MP	DFS PA VC	3	M 1-3
	6.2	Implementierung des sensorischen Equipments in Abstimmung mit der DFS	MP	DFS VC	2	M 2-5
	6.3	Initialerhebung der physiologisch / psychologischen IST-Zustände der Lotse	MP	DFS VC	3	M 6-8
	6.4	Datenintegration (IST-Zustände - Abgleich der Pilotenlogfiles)	MP	VC	1	M 9
	6.5	Konzeption des abstrakten Modells der möglichen Zustände von Emotioner	MP	PA KI	3	M 10-12
	18.1	Prototypische, parallele Evaluationsdurchführungen	MP	alle	6	M 25-30
	18.2	Implementierung von neuen Erkenntnissen in das Modell	MP	alle	6	M 31-36
Mimik	4.1	Extraktion von Mimik mittels Active Appearance Modellen	KI	MP	8	M 1-8
	4.2	Entwicklung eines Deep-Learning Active Appearance Modells zur Extraktion	KI	keine	8	M 9-16
	10.1	FACS Codierung von Gesichtsausdrücken	KI	MP	4	M 17-20
	10.2	GPU basiertes Modell zur Extraktion von Mimik-Merkmalen	KI	MP	4	M 21-24
	14.1	Optimiertes Modell zur FACS Codierung von Mimik-Merkmalen	KI	GDV	6	M 25-30
	14.2	Modell zur Detektion zur Erfassung der zeitlichen Entwicklung von Gesichtsa	KI	keine	6	M 31-36
Gestik	1	Aufbau des sensorischen Systems in abstrahierter Arbeitsumgebung	GDV	alle	3	M 1-3
	3.1	Auswertung der stereoskopischen Bilddaten	GDV	PA MI	6	M 4-9
	3.2	Rekonstruktion der Körperoberfläche	GDV	keine	3	M 10-12
	9.1	Definition des Skelettmodells	GDV	PA	3	M 13-15
	9.2	Erzeugung von Bewegungsdaten der Teammitglieder	GDV	keine	6	M 16-21
	9.3	Extraktion der Blickrichtung	GDV	KI	3	M 22-24
	13.1	Definition von relevanten Referenzposen und Extraktion der entsprechende	GDV	keine	3	M 25-27
	13.2	Analyse der Bewegungsdaten	GDV	keine	3	M 28-30
	13.3	Definition von relevanten Referenzgesten und Extraktion der entsprechende	GDV	keine	3	M 31-33
	13.4	Analyse der Bewegungsdaten in Bezug auf Gesten	GDV	keine	3	M 34-36
Audioanalyse	5.1	Konzeption Evaluationskorpus und Erfassung von Audiodaten	MI	alle	4	M 1-4
	5.2	Transkription für den Evaluationskorpus	MI	MP	3	M 5-7
	5.3	Erstellung von Metadaten für den Evaluationskorpus	MI	MP	5	M 8-12
	11.1	Feature-Extraktion	MI	MP PA KI	6	M 13-18
	11.2	Linguistische Merkmale	MI	MP PA	6	M 19-24
	15.1	Spracherkennung	MI	MP	6	M 25-30
	15.2	Abgleich und Integration	MI	PA	6	M 31-36
Benutzungsoberfläche	7.1	Analyse aktueller Nutzerschnittstellen	VC	MP, DFS	3	M 1-3
	7.2	Bedien- und Visualisierungskonzepte	VC	MP PA DFS	4	M 4-7
	7.3	Design der adaptiven Nutzerschnittstelle	VC	MP PA DFS	5	M 8-12
	12.1	Visualisierungsanalyse Emotionsmodell	VC	MP	3	M 13 - 15
	12.2	Interaktions- und Visualisierungsdesign Emotionsmodell	VC	alle	3	M 16 - 18
	12.3	Visualisierungsprogramm Emotionsmodell	VC	MP	6	M 19 - 24
	17.1	Implementierung adaptive Nutzerschnittstelle	VC	MP PA	6	M 25 - 31
	17.2	Evaluierung adaptive Nutzerschnittstelle	VC	MP	6	M 32 - 36

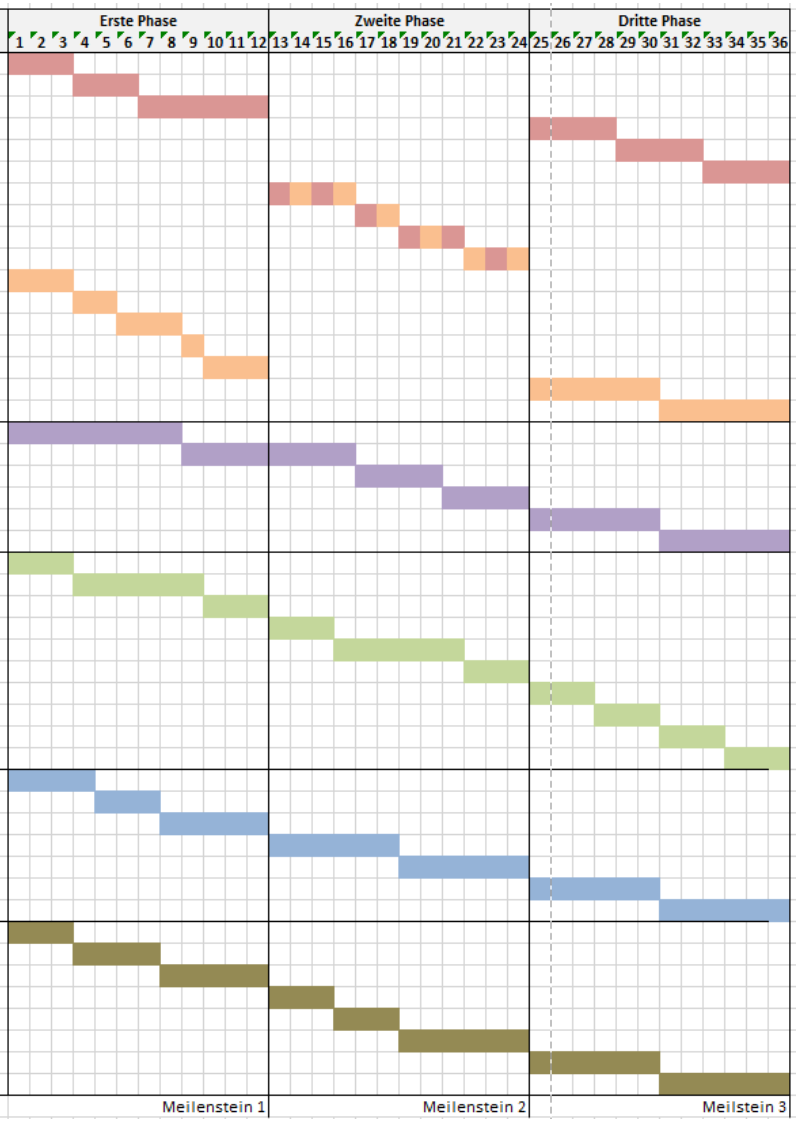


Abbildung 2: Arbeitsplan

Im September 2017 wurde beim Projektträger eine kostenneutrale Verlängerung der Projektlaufzeit bis zum 30. April 2018 beantragt, die mit dem Schreiben vom 28. November 2017 bewilligt wurde. Die Verlängerung wurde beantragt, da eine unerwartete Schwierigkeit bei der Erstellung des Emotions- und Kommunikationsmodells aufgetreten war. Wie im Arbeitsplan vorgesehen, wurden mögliche Einflussfaktoren hinsichtlich ihrer statistischen Relevanz in den Datensätzen untersucht, die in den betriebsinternen Simulationen der DFS aufgezeichnet wurden. Allerdings wurde die Modellbildung dadurch erschwert, dass die Daten nur geringe Varianzen in den psychophysiologischen Belastungswerten aufweisen (siehe 2. Sachbericht S.18). Um dennoch ein Modell erstellen zu können, wurde von uns eine eigene Simulation entwickelt, die die Fluglotsenarbeit vereinfacht nachstellt und eine Induktion hoher Belastungszustände bei Variation der emotionalen Befindlichkeit ermöglicht. Dieses Experiment wurde zunächst mit studentischen Probanden durchgeführt und lieferte verwertbare Daten zur Modellbildung. Um das Experiment auch mit Fluglotsen durchführen und auswerten zu können, war eine längere Projektlaufzeit notwendig, da die Terminplanung der Studie mit Fluglotsen eine längere Zeit in Anspruch nahm. Die finale Messung wurde im Dezember 2017 durchgeführt. Die dabei erhobenen Daten wurden in 2018 ausgewertet.

Für das Projekt wurden drei Meilensteine definiert, um die erzielten Projektfortschritte zu präsentieren und kritisch zu reflektieren. Der erste Meilenstein bestand in der Präsentation der entwickelten Konzepte zur Datenerfassung und des Kommunikationsmodells, auf einem öffentlichen Workshop mit Wissenschaftlern und potentiellen Anwendern. Der Meilenstein wurde durch die Präsentation von Dr. Nicholas Müller im November 2015 auf der Tagung der Deutschen Gesellschaft für Luft- und Raumfahrt (DGLR) erfüllt [113]. Neben dem Fachpublikum der Tagung waren Herr Buxbaum, der der Ansprechpartner des Projektes bei der DFS war, sowie einige seiner Kollegen anwesend. Darüber hinaus waren potentielle Anwender wie das Deutsche Zentrum für Luft- und Raumfahrt (DLR), das Fraunhofer-Institut für Kommunikation, Informationsverarbeitung und Ergonomie (FKIE) sowie Automobilbauer vertreten.
Die zweite Projektphase schloss mit einem öffentlichen Workshop am 09. März 2017 an der TU Chemnitz ab, auf dem der Projektstand und die Leistungsfähigkeit der entwickelten Algorithmen der interessierten Öffentlichkeit vorgestellt wurden.

Der dritte Meilenstein wurde mit der Vorstellung der Projektergebnisse auf dem Vernetzungstreffen des BMBF-Förderschwerpunktes „Sozial- und emotionssensitive Systeme“ (InterEmotio) erfüllt, welches vom 31. Januar bis 01. Februar 2018 in Bonn stattfand. Prof. Dr. Brunnett stellte die erzielten Projektergebnisse in einem Vortrag den Teilnehmerinnen und Teilnehmern vor. Ein Projektstand mit Demonstratoren und Informationsmaterial, der von Projektmitarbeitern betreut wurde, bot den Teilnehmerinnen und Teilnehmer die Möglichkeit, sich über das Projekt und die erzielten Ergebnisse auszutauschen.

Des Weiteren wurden während der Projektlaufzeit zwei Workshops zu ELSI-Fragestellungen (ethische, rechtliche und soziale Implikationen des Projektes) durchgeführt. Der erste Workshop wurde am 10. November 2016 zusammen mit der DFS und in Abstimmung mit dem Betriebsrat in der DFS Niederlassung Bremen durchgeführt. In dem Workshop informierten die Projektmitarbeiterinnen und –mitarbeiter die teilnehmenden Fluglotsen über die bereits gewonnenen Erkenntnisse. Auf dieser Grundlage diskutierten die Teilnehmerinnen und Teilnehmer ELSI-Aspekte der Forschung. Der Workshop diente somit zur Exploration der Nutzerakzeptanz und zur Diskussion der möglichen Auswirkungen des MACe-Lot-Ansatzes auf die Arbeit der Fluglotsen. Der Workshop wurde von den Teilnehmenden als Erfolg gewertet.

Der zweite ELSI-Workshop wurde im Rahmen der Konferenz Mensch und Computer 2017 in Regensburg von Prof. Eibl, Prof. Ohler, Dr. Valtin und Dr. Müller organisiert. Der Workshop mit dem Titel „Ethische Herausforderungen in sozioinformatorischen Forschungsprojekten“ wurde mit einem Call for Papers beworben. Mit insgesamt 15 Teilnehmerinnen und Teilnehmern war der Workshop sehr gut besucht. Anhand konkreter Projekte wurden ELSI-relevante Problematiken, entsprechende Lösungsansätze und Handlungsempfehlungen diskutiert.

4. Wissenschaftlicher und technischer Stand, an den angeknüpft wurde

Aus **kognitionspsychologischer Sicht** ist die Arbeit von Fluglotsen ein äußerst ergiebiges Thema. Die mentale Repräsentation des Problemraums, von den Lotsen selbst „Picture“ genannt, die unterschiedlichen Strategien und Heuristiken der Lotsen, das Zusammenspiel interner und externer Repräsentationen (Medien) beim Arbeitsprozess, die Einführung neuer Technologien (z. B. stereoskopische 3D Anzeigen; Assistenz- und Entscheidungsunterstützungssysteme), die Entwicklung von Lernprogrammen (z. B. Intelligente Tutorielle Systeme) für Lotsenschüler sind Beispiele für kognitive Themen. Im Zentrum der Forschung steht die Auslastung des Arbeitsgedächtnisses der Lotsen bei ihrer Arbeit, durch welche Parameter diese beeinflusst wird und wie man sie reliabel und valide messen kann [74], sowie die Konzepte „Stress“ und „Coping“ (Stressbewältigung). Neben subjektiven Maßen (Online-Fragebögen) wird versucht den Workload mit physiologischen Maßen zu erfassen (z.B. [164], [19]). Das Zusammenspiel mehrerer Lotsen wird versucht mit Modellen aus dem Bereich „distributed cognition“ zu konzipieren [46].

Im Projektverlauf wurden darüber hinaus noch weitere empirische Forschungsbeiträge in den Korpus übernommen. Darunter ist beispielhaft [166] zu erwähnen, in welchem eine Automatisierung von Fluglotsenarbeitsplätzen negative Auswirkungen auf die kognitive Leistung zeigen könnte. Ebenso zu nennen sind neuere Untersuchungen zur kognitiven Belastung [31], [104], insbesondere bei Fluglotsen [124].

Die **Vorhersage und Berechnung des mentalen Zustandes** von Menschen im Allgemeinen ([84],[155], [138]) und speziell von Fluglotsen [74] ist ein aktuelles Forschungsthema.“ Allerdings ist es noch ein langer Weg zu einem kompletten Verständnis von mentalem Workload“ [74] und der damit verbundenen robusten Vorhersage des organisatorischen Entscheidungsverhaltens von Menschen insbesondere Fluglotsen. Einzelne Arbeiten versuchen das Zusammenspiel (meta)kognitiver und emotionaler Komponenten mittels intelligenter Agentensysteme zu modellieren [47]. Eine Zusammenführung mehrerer Eingabemethoden, wie Sprache, Bild und physiologische Messungen wurden bisher noch nicht angewendet. Hierbei kann das vorgestellte Projekt einen wesentlichen Beitrag zur neuesten Forschungsentwicklung leisten. Dabei werden die Erfolgsaussichten durch die Möglichkeit der Integration von Entwicklungsstufen des Systems in den Simulationskontext der DFS, die regelmäßig Systemvarianten experimentell variiert und den Einfluss auf psychische Determinanten des Arbeitsprozesses erhebt, deutlich erhöht.

Dazu wurden im Projektverlauf ebenfalls weitere Literaturquellen in den Korpus der Arbeit übernommen. Darunter fallen beispielsweise die Untersuchungen von [20], sowie zur adaptiven Aufgabenzuweisung [134] und psychophysiologischen Veränderungen [17].

Zur **Erfassung menschlicher Bewegungen** (Motion Capturing) existieren verschiedene Techniken (Überblick siehe [94]), wobei spezielle Hardware, z.B. Datenhandschuhe und Markertracking eine hohe

Genauigkeit und geringe Latenz ermöglicht, aber auch zur Beeinflussung menschlicher Bewegungen und zu geringer Nutzerakzeptanz führt (siehe [33]). Deshalb ist die kamerabasierte Vermessung Gegenstand der aktuellen Forschung, wobei das Ziel darin besteht, in realen Umgebungen ohne zusätzliche technische Hilfsmittel auszukommen. Derartige Verfahren für den Nahbereich, die z.B. in Spielkonsolen Einsatz finden, implementieren so genannte structured-light (Überblick in [118]) bzw. time-of-flight Techniken [122]. Dabei wird Infrarotlicht projiziert und seine Ausbreitung fotografisch ausgewertet. Prinzip-bedingt eignen sich diese Verfahren nicht für helle, große Räume. Silhouetten basierte Techniken [18] überwinden diese Nachteile, setzen jedoch mehrere Perspektiven und einen guten Kontrast zwischen Person und Hintergrund voraus. Kommerzielle Echtzeitverfahren, z.B. Organic Motion, benötigen deshalb monochromatische Hintergründe. Stereoskopische Kamerasysteme können räumliche Bewegungen ähnlich zu den IR Verfahren aus einem einzigen Blickwinkel erfassen, wobei die Kombination mehrerer Kameraperspektiven zum Erreichen stabiler Daten prinzipiell möglich ist [6]. Diese Technik bietet das größte Potential zur Bewegungserfassung in realen Arbeitssituationen und soll im Projekt zum Einsatz kommen. Die **Erkennung von Gesten** erfolgt in einem zweistufigen Prozess aus Bewegungserfassung und anschließender Aktionsdetektion über eine Mustererkennung anhand von zuvor erfassten Referenzbewegungen bzw. durch direkte Einpassung eines Skelettmodells (ein Überblick in [133]). Letzteres ermöglicht prinzipiell eine höhere Flexibilität, wobei entsprechende Vorarbeiten an der Professur GDV durchgeführt wurden.

Die Verarbeitung von Videodaten, Extraktion und Klassifikation der Pose und der Bewegung sind Schwerpunkte der aktuellen Forschung in den Bereichen der Informatik, Informationstechnik und Künstlichen Intelligenz sowie Arbeit- und Sozialwissenschaften. Daher sind eine Vielzahl neuer Erkenntnisse und fertig einsetzbarer Algorithmen während der gesamten bisherigen Projektlaufzeit veröffentlicht worden. Zur Erkennung der **Körperpose** sind mittlerweile vielversprechende Algorithmen verfügbar [154][176]. Zur Erfassung der **Kopfpose** existieren ebenso neue Arbeiten [96][116]. Aus der zeitlichen Abfolge der Bilder können zusätzliche Informationen gewonnen werden, die neben der besseren Erkennung der **Pose** auch eine Klassifikation der **Bewegung** erlauben [25].

In den letzten Jahren gab es viele Fortschritte hinsichtlich intelligenter Systeme, die automatisch **emotionale Gesichtsausdrücke** in Bildern oder Videosequenzen auswerten. Allerdings konzentrieren sich viele Ansätze auf die Unterscheidung von wenigen emotionalen Ausdrücken (Freude, Trauer, Angst, Überraschung und Ekel), wie beispielsweise durch Ekman [29] definiert. Tatsächlich sind menschliche Gesichtsausdrücke deutlich vielfältiger. Diese große Vielfalt kann besser durch eine low-level Kodierung in Form des Facial Action Coding Systems (FACS) erfasst und beschrieben werden [36]. Demnach sind Gesichtsausdrücke durch 44 Action Units beschrieben, die sich auf bestimmte Gesichtsmuskeln beziehen. Wenngleich es verschiedene Ansätze zur Extraktion von Merkmalen zur Bestimmung der Action Units gibt, haben sich sogenannte Active Appearance Models ([20] [82]) als sehr interessant erwiesen. Active Appearance Modelle sind ein statisch basierter Template Matching Ansatz, der sowohl die Grauwertverteilung (lokale Feature) als auch die Gesichtsform (global Features) berücksichtigt. Durch die iterative Kombination von lokaler und globaler Optimierung sind Active Appearance Modelle allerdings sehr aufwändig zu berechnen und können Konvergenzprobleme aufweisen, wenn sie nicht richtig initialisiert sind. Da sich Gesichtsausdrücke dynamisch entwickeln, sollten diese auch in Videosequenzen analysiert und detektiert werden, wobei dieses nicht lediglich auf voneinander unabhängigen Einzelbildern erfolgen solle. Erste Ansätze nutzen beispielsweise optische Fluss-Information zum robusten Tracking von lokalen Active Appearance Merkmalen [26].

Die klassische **Audioanalyse zur Emotionserkennung** basiert auf einer Analyse der low-level-Features (für einen Überblick siehe [33]). Dabei werden spektrale Analysen wie zum Beispiel Mel-Frequenz-Cepstrum-Koeffizienten (MFCC) verwendet, die sowohl bei gesprochener Sprache als auch bei Musik eingesetzt werden können. Daneben werden speziell für Sprachanwendungen auch sehr vielversprechend prosodische Besonderheiten (typischerweise: Mean pitch, Mean energy, Pitch variance, Skew of logarithmic pitch, Range of logarithmic pitch, Range of logarithmic energy) untersucht (z. B. [86]). Die Methoden zur Extraktion akustischer sowie deren abgeleiteten Merkmale stehen zum Teil bereits als Open Source-Lösungen wie OpenSMILE [34] zur Verfügung. Desweiteren existiert eine Vielzahl an Klassifikationsverfahren zur Schätzung von Emotionen (vgl. [7]). Im Gegensatz zur automatischen Spracherkennung mit Hidden-Markov-Modellen lässt sich jedoch keine allgemeine Aussage über den Einsatz eines besonders geeigneten Klassifikationsverfahrens treffen. Hinsichtlich der Robustheit sind Support Vektor Maschinen besonders hervorzuheben (z. B. [113]).

Von zentraler Bedeutung für die Qualität dieser Verfahren ist ein geeigneter Datenbestand, der für das Training und die Evaluation verwendet werden kann. Hier stehen nur wenige und relativ kleine Datenbestände zur Verfügung wie zum Beispiel die deutsche EMO-DB [17] (10 Sprecher mit je 10 Äußerungen in den Emotionen fröhlich, ärgerlich, ängstlich, gelangweilt, angeekelt). Einen vergleichenden Überblick gibt [125]. Der hier vorgeschlagene Anwendungsfall ist jedoch zu speziell, als dass diese Daten geeignet wären. Zum einen wird von den Centerlotsen ein sehr spezieller Wortschatz verwendet, der hier auch Berücksichtigung finden muss. Zum anderen sind die Centerlotsen dahingehend ausgewählt und trainiert, mit emotional herausfordernden Situationen umzugehen. Eine emotionale Erregung wird daher deutlich schwieriger zu messen sein als es in Standardsituationen der Fall ist. Im vorgeschlagenen Projekt soll zunächst nur die negative Erregung identifiziert werden und keine speziellen Ausprägungen wie Ärger, Angst oder Ekel. Dabei wird ein trainiertes Regressionsmodell zur Messung von Arousel und Valence erstellt.

Linguistische Merkmale des Gesprochenen tragen ebenso zur Emotionserkennung bei (z. B. [66][118]). Dies spiegelt sich in emotions-gewichteten Wörtern sowie Grammatiken bis hin zu höheren Semantiken wider. In diesem Zusammenhang gibt es eine Vielzahl unterschiedlicher Ansätze und Methoden zur Verarbeitung textueller Daten, wobei Class-based N-Grams und Vector Space Modeling sehr vielversprechend sind. Einen Überblick gibt [7]. Diese statistischen Verfahren sind gut geeignet, um untypische Äußerung, bzw. Äußerungen in untypischen grammatikalischen Zusammenhängen schnell zu identifizieren. Solche Auffälligkeiten können gerade in verbal recht eng umgrenzten Szenarien wie der Kommunikation von Fluglotsen genutzt werden, um besondere Situationen zu erkennen.

Die **Sensordatenfusion** ist ein typisches Problem in vielen Bereichen der Technik, bei denen redundante Sensordaten aus verschiedenen Quellen mit unterschiedlicher Qualität (Kovarianz, Ausreißer) und verschiedenen Messfrequenzen zu einer verbesserten Gesamtaussage in Form eines Zustandsvektors verknüpft werden müssen. Die Sensordatenfusion kann auf unterschiedlichen Leveln erfolgen: Auf dem Merkmalsebenenlevel (Frühe Fusion) [182] oder dem Entscheidungslevel (späteren Fusion) [183]. Um dem Problem entgegenzuwirken, dass alle Eingabekanäle und deren Merkmalserkennungen verschiedene Modalitäten haben und diese schwer untereinander abgewogen werden können, wird im Folgenden eine Fusion auf Entscheidungslevel angestrebt. Des Weiteren werden Korrelationen zwischen den verschiedenen Modalitäten der Eingabekanäle ermittelt, welche dann die Kovarianzen des Fusionsvektors darstellen [184].

Die **visuelle Darstellung des Zustandsmodells** des Interaktionsverhaltens und der zugrundeliegenden emotionalen Hinweisreize ist sowohl für die Entwicklung des Gesamtsystems als auch für die Anwendung des Systems für einen Supervisor unverzichtbar. Die Visualisierung von multidimensionalen zeitveränderlichen Daten [3] ist ein weites und komplexes Feld im Bereich der Informationsvisualisierung [132]. Für fast jeden möglichen Datentyp und jede Visualisierungsaufgabe [117] gibt es mittlerweile mehrere Ansätze wie die Daten darzustellen sind. Jeder der bekannten Ansätze beansprucht verschiedene Stärken bezüglich der Daten und Aufgaben. Allerdings sind dies nur die Werkzeuge, um die eigentliche Aufgabe des Designs, der Implementierung und der Evaluation eines Visualisierungssystems für einen ganz speziellen Anwendungsfall zu lösen. Dieses Feld der menschen- und aufgabenorientierten Visualisierung [41] hat besonders in den letzten Jahren immer mehr an Bedeutung gewonnen. Nur durch genaue Analyse der darzustellenden Daten, der zu lösenden Aufgaben und der angesprochenen Nutzergruppe ist es möglich, eine Visualisierungsumgebung zu schaffen die sich als sinnvoll und effizient erweist. Zum aktuellen Zeitpunkt gibt es kein Visualisierungssystem, welches in diesem Kontext dazu in der Lage wäre.

Um die Emotionserkennung schließlich in **Rückkopplung für die Bedienoberflächen**, die den einzelnen Fluglotsen zur Verfügung stehen, nutzbar zu machen, bietet sich das Konzept der Adaptiven Benutzungsoberflächen (Adaptive User Interface, AUI). Emotionserkennung hat hier in den letzten Jahren vor allem in sprachbasierten Dialogsystemen und Entertainmentsystemen Einzug gehalten. Die Zielsetzung ist dabei klar definiert: Verärgerte Anrufer und einschlafende Autofahrer. Die Erfahrungen mit AUIs sind dabei ambivalent. So können AUIs tatsächlich Prozesse sehr beschleunigen, sie können aber auch im Gegenteil die Bedienung radikal erschweren - schließlich müssen sie ganz grundsätzlich gegen das Konsistenzgebot der Mensch-Computer-Interaktion verstoßen [79]. Es kristallisiert sich heraus, dass AUIs sehr speziell für den einzelnen Anwendungsfall konzipiert werden müssen. Hier gilt es den Bereich der Adaption etwa in Hinblick auf Navigationsstruktur und Visualisierung wie auch den Umsetzungsgrad - aggressiv bis sehr zurückhaltend - genau zu evaluieren. Fluglotsensysteme sind dabei wissenschaftlich hochinteressant, da der Anwendungskontext extrem sensibel und sicherheitskritisch ist. Eine Lösung in diesem Anwendungskontext lässt die Chance zu, AUI-Pattern für sicherheitskritische Interaktionen zu definieren, die auch in anderen Kontexten Bestand haben können.

Im Bereich der adaptiven Automatisierung für Fluglotsen wurden bereits vielversprechende Ergebnisse in Bezug auf Adaption auf Grundlage der Arbeitslast erlangt [67][108][135]. Diese Ergebnisse betrachten allerdings lediglich den Automatisierungsgrad und weniger visuelle und strukturelle Anpassungen der Nutzeroberfläche. Zudem wurde die Entwicklung des Interaktions- und Visualisierungskonzeptes auch auf andere Forschungsarbeiten zu den Nutzungsschnittstellen von Fluglotsen beeinflusst. Dies beinhaltet beispielsweise Arbeiten zur Dimensionalität der Bildschirmanzeige [8][148][175], zu Höhen- [105][151] und Wettervisualisierungen [30], oder zur Arbeitsorganisation mit Flugstreifen [56][165][174]. Für eine Übersicht über Interaktions- und Visualisierungsmethoden siehe auch [113].

5. Zusammenarbeit mit anderen Stellen

Während des Projektes fand eine enge Zusammenarbeit der Verbundpartner mit der Deutschen Flugsicherung GmbH statt. Im November 2015, Mai und September 2017 sowie im Dezember 2017 fanden Datenerhebungen bei der DFS statt. Ein Vertreter des DFS nahm zudem an den beiden öffentlichen Statustreffen des Projektes am 16. Juli 2015 und am 09. März 2017 teil. Am 10. November 2016 fand

in der DFS Niederlassung Bremen ein Workshop mit Fluglotsen und Projektmitarbeitern statt, in welchem die ethischen, rechtlichen und sozialen Implikationen (ELSI) des Projektes diskutiert wurden. Neben der thematischen Auseinandersetzung zu ELSI-Aspekten bot der Workshop eine gute Möglichkeit für den Austausch zwischen den Fluglotsen und wissenschaftlichen Mitarbeiterinnen und Mitarbeitern. Weitere Einzelheiten der Zusammenarbeit werden unter Abschnitt II.1 („Verwendung der Zuwendung und des erzielten Ergebnisses im Einzelnen“) in den betreffenden Arbeitspaketen beschrieben.

II. Eingehende Darstellung

1. Verwendung der Zuwendung und Ergebnisse des Projektes

Die Darstellung der Ergebnisse erfolgt entsprechend den zentralen Teilbereichen des Projektes. Die Zuordnung zu den 18 Arbeitspaketen ist kenntlich gemacht. Die Zuwendungen wurden entsprechend der Projektzielsetzung zur Bearbeitung dieser Arbeitspakete eingesetzt.

1.1 Bewegungsanalyse (AP 1, AP 3, AP 9, AP 13)

Aufbau des sensorischen Systems in abstrahierter Arbeitsumgebung

Nach einer Besichtigung der Arbeitsweise der Fluglotsen in der realen Situation im Center in Langen, einer Simulation der Fluglotsentätigkeit im Rahmen der Fluglotsenausbildung in der Akademie in Langen sowie eines Besuchs der Simulationsumgebung im Forschungszentrum in Langen erfolgte die Planung des sensorischen Systems anhand der vorliegenden Arbeitsumgebung.

Basierend auf vor Ort vorgenommenen Messungen und Unterlagen der DFS über die Abmessungen eines Simulationsarbeitsplatzes wurde ein dreidimensionales Modell des Arbeitsplatzes erstellt. Dieses umfasst drei Fluglotsenspulte sowie eine vereinfachte Darstellung des umgebenden Raumes. Auf Grundlage des dreidimensionalen Modells wurde die Position, Ausrichtung und Anzahl der Kameras und der benötigten Mikrofone sowie der Platzierung der Rechentechnik geplant.

Im nächsten Schritt erfolgte ein den Abmessungen der Simulationsarbeitsplätze entsprechender attrappenhafter realer Nachbau zweier benachbarter Fluglotsenpulte aus Aluminiumprofilen. Auf eine Verkleidung der technischen Ausstattung wurde verzichtet, um die Mobilität und Adaptierbarkeit des Prototyps zu gewährleisten.

Anhand des realen Nachbaus wurden Testmessungen mit verschiedenen Sensoren (insbesondere Kameras verschiedenen Typs) durchgeführt, um die optimale Positionierung des Messequipments am Arbeitsplatz zu bestimmen.

Aufgrund der räumlichen Situation der Arbeitsplätze der Centerlotsen ist es nicht möglich, die Fluglotsen per Tiefenkamera von der Seite oder von hinten zu erfassen, ohne die Arbeitsabläufe zu stören. Aus diesem Grund kann lediglich eine Erfassung der dem Arbeitsplatz zugewandten Seite erfolgen. Aufgrund der Nähe der Arbeitsplätze der Teammitglieder ist es ausreichend, eine Tiefenkamera pro Fluglotse zu verwenden, wobei der Bereich in der Mitte beider Arbeitsplätze von beiden Kameras erfasst wird. Auf diese Art kann sichergestellt werden, dass bei Interaktion der Teammitglieder bis hin zu dem Szenario, dass beide Lotsen sich am gleichen Arbeitsplatz befinden, zu jedem Zeitpunkt beide Lotsen erfasst werden können.

In zwei Testmessungen an den Simulationsarbeitsplätzen in der Forschungsabteilung der DFS wurde das Messsystem dahingehend validiert, dass es für die Aufnahme der benötigten Daten an den derzeit vorhandenen Arbeitsplätzen in Langen geeignet ist. Somit ist der Einsatz des nachgebauten Arbeitsplatzes für Datenerhebung, Evaluation und Demonstration an der TU Chemnitz möglich.

Auswertung der stereoskopischen Bilddaten

Die Aufgabe bestand zunächst darin, die Bildpunkte einer Kamera dem Menschen bzw. der Umgebung zuzuordnen. Deshalb fand eine Recherche zu Verfahren statt, welche Menschen in Farb- oder Tiefenbildern lokalisieren können und idealerweise gleichzeitig ausgeben, welche Bildpunkte dem Menschen zugeordnet sind. Nach der Evaluation mehrerer Algorithmen, stellte sich der von [146] zunächst als besonders geeignet heraus.

Dieses Verfahren nutzt ausschließlich die Tiefenbilder und ist daher robust gegenüber Änderungen der Farbe der Kleidung. Er liefert nicht nur Auskunft darüber, welcher Bildpunkt dem Menschen zugeordnet ist, sondern zusätzlich noch die Angaben, um welche Körperpartie es sich handelt, z. B. linke oder rechte Gesichtshälfte, Rumpf oder Schultern. Dadurch wird die nachfolgende Einordnung der Skelettelemente erleichtert. Es stellte sich jedoch heraus, dass für die korrekte Funktion des Verfahrens sehr genaue Tiefenbilder notwendig sind. Mehrere Messungen im Labor und am Arbeitsplatz der Fluglotsen in Langen zeigten einen erheblichen Einfluss der Beleuchtung und Sitzposition des Lotsen auf die Qualität der Tiefenbilder. Es stellte sich als aufwendig heraus die Kamera so auszurichten, dass der Algorithmus kontinuierlich gut funktioniert. Die Fehlertoleranz des Algorithmus war damit für den geplanten Einsatz als zu gering zu bewerten.

Die Recherche wurde daher mit geänderter Strategie fortgesetzt. Das finale Ziel der Bildauswertung ist die Körperhaltung einer Person im 3D-Raum digital zu erfassen. D. h. eine genaue Zuordnung der Bildpunkte zum Menschen und eine Rekonstruktion der Körperoberfläche sind nur Zwischenschritte, die letztlich nicht mehr benötigt werden. Es reicht daher vollkommen aus, die Position einzelner Körperteile anhand markanter Merkmale zu messen und sie logisch der korrekten Person zuzuordnen. Die Hinzufügung einer Tiefeninformation geschieht dann nicht mehr in Form einer detaillierten Oberfläche, sondern als einzelner Wert je Körperpartie.

Es wurden daher mehrere Methoden untersucht, welche direkt aus dem Farbbild heraus Menschen erkennen und lokalisieren können. Dabei zeichnete sich [128] als besonders zuverlässig aus. Im nächsten Schritt kann bereits ein 2D-Skelett in den erkannten Menschen eingepasst werden. Daher wurde weiter nach Verfahren gesucht, welche die Körperpartien im 2D-Farbbild unter Beachtung menschlicher Proportionen finden und somit unmittelbar das 2D-Skelett wiedergeben. Bei der Evaluation dieser Verfahren stellte sich [21] als besonders genau heraus. Unter Einsatz dieser Methode sind wir jetzt in der Lage 2D-Skelette aus Farbbildern zu extrahieren.

Damit waren alle Voraussetzungen erfüllt, um das 3D-Skelett erfassen zu können.

Definition des Skelettmodells

Die Anforderungen an das Skelettmodel ergeben sich aus der Aufgabe, die arbeitsbedingte Interaktion mit den Piloten und den Kollegen im Kontext von emotionaler, psychomentaler und psychosozialer Beanspruchung bzw. Belastung zu messen. Weitere Hinweise auf die Beanspruchung könnten aus nicht interaktionsbezogenen Handlungen ablesbar sein.

Relevante Daten für die Messung der Beanspruchung sind unter anderem aus der Rumpfhaltung zu erwarten. Konzentriertes Arbeiten könnte mit vorgebeugter Haltung korrelieren. Die Position des Kopfes und der Schultern ermöglichen Aussagen über die Interaktion innerhalb der Dyade. Armbewegungen deuten auf die Verwendung von Eingabegeräten wie zum Beispiel der Maus oder des Touchscreens für die Flugstreifen hin.

Die Repräsentation des menschlichen Skeletts im Motion-Capture Kontext erfolgt durch eine Kombination eines Satzes von Gelenkwinkeln mit Knochen einer definierten Länge. Eine Bewegung wird durch die Veränderungen der Gelenkwinkel dargestellt.

Die Verwendung von Gelenkwinkeln gewährleistet die Invarianz gegenüber variierender Körpergröße sowie Vergleichbarkeit zwischen verschiedenen Personen. Die in der Computergrafik verwendeten Repräsentationen des menschlichen Skelettes umfassen in der Regel folgende Gelenke: fünf Gelenke, welche die Rückenwirbelsäule abstrahieren; zwei Gelenke der Halswirbelsäule; je ein Schulter-, Ellbogen- sowie Handgelenk pro Arm sowie ein Hüft-, Knie- und Fußgelenk pro Bein.

Bedingt durch die sitzende Tätigkeit der Fluglotsen an einem Tisch sowie die Anbringung der Stereokamera ist lediglich eine Erfassung des Oberkörpers möglich, sodass das verwendete Skelett auf die Gelenke der Beine verzichten kann.

Die genaue Erfassung aller Gelenkwinkel des Skelettmodells würde eine Vielzahl an optischen Markern sowie die Verwendung mehrere Kameraperspektiven erfordern. Eine derartige Datenerfassung ist jedoch sehr invasiv und nur unter kontrollierten Laborbedingungen möglich, da für die korrekte Datenerfassung eine konstante Beleuchtungssituation gegeben sein muss und eine zeitintensive Kalibrierung des Systems nötig ist.

Weder die Umgebung der Simulation in der Forschungsabteilung der DFS in Langen, noch die Arbeitsplätze im realen Center würden einen Einsatz mehrerer Kameras erlauben, ohne die Arbeit selbst zu stören. Die Verwendung jeglicher am Körper angebrachter Marker stellt durch damit verbundene mögliche Ablenkungen oder Einschränkungen der Bewegungsfreiheit ein erhebliches Sicherheitsrisiko dar und ist somit ebenso nicht möglich.

In der Simulationsumgebung in Langen hat sich der Einsatz einer minimalinvasiven Frontalkamera als praktikabel erwiesen, welche oberhalb des Radarschirms befestigt ist und sich damit außerhalb des Blickfeldes des Lotsen befindet.

Die durch diese Kamera gelieferten Daten werden verwendet, um ein perspektivabhängiges 2D-Skellett des Lotsen zu erfassen. Auf Basis der biomechanischen Beschränkungen des menschlichen Skeletts sowie Kenntnis über Position und Eigenschaften der Kamera, ist es nun möglich, aus den 2D-Skeletten durch Abgleich mit einer Datenbank von 3D-Skeletten der eingenommenen Posen eine Umwandlung der Skelette in den dreidimensionalen Raum vorzunehmen.

Da es sich bei der Arbeit der Centerlotsen um eine sitzende Tätigkeit handelt, ist der Zustandsraum an zu erwartenden Posen stark eingeschränkt. Die Position und Ausrichtung der Lotsen gegenüber dem Arbeitsplatz sind durch die feststehende Einrichtung handlungsbezogen festgelegt. Das 2D-Skelett ist perspektivabhängig, d. h. die gleiche Körperhaltung ergibt unterschiedliche Skelette bei Drehung des gesamten Körpers. Eine derartige Veränderung der Pose ist damit bereits anhand der Skelettänderung erkennbar. Ein Vergleich von Posen unterschiedlicher Orientierungen zum Arbeitsplatz wäre mit einem 3D-Skelett konsistenter möglich. Dies ist jedoch für unser Projekt nicht relevant, da unterschiedliche

Sitzrichtungen mit unterschiedlichen Arbeitssituationen korrelieren, wobei zwangsläufig eine andere Haltung eingenommen wird. Ein Hindrehen zum Kollegen macht beispielsweise die Arbeit am Radarbildschirm unmöglich.

Die Gelenkhierarchie wird durch den Wechsel vom zweidimensionalen in den dreidimensionalen Raum nicht verändert, einzig die Werte der Gelenkwinkel unterscheiden sich. Eine Änderung des Skelettmodells ist nicht erforderlich.

Erzeugung der Bewegungsdaten der Teammitglieder

Im April 2017 wurde eine frei verfügbare C++ Bibilothek names *OpenPose* [103] veröffentlicht, welche die gleichzeitige bildbasierte Skeletterstellung mehrerer Personen im Farbbild einer Kamera ermöglicht. Die bisherigen Verfahren der Skeletterkennung *Human Pose Estimation with Iterative Error Feedback (IEF)* [21] *und Convolutional Pose Machines* [170] hatten demgegenüber den Nachteil jeweils nur ein Skelett erstellen zu können. Um das Ziel der simultanen Bewegungsauswertung beider Fluglotsen einer Dyade gewährleisten zu können, wurde bisher stets eine stereoskopische Kamera (*Multisens S7)* pro Fluglotse verwendet. Der Einsatz von OpenPose gestattet die Skeletterkennung beider Fluglotsen mit derselben Kamera durchzuführen, sofern die Arbeitsplätze direkt nebeneinanderliegen. Damit kann die Erfassung beider Fluglotsen mit beiden Kameras erfolgen, um mehrere Perspektiven zur Erstellung von 3D Skeletten nutzen zu können. Dabei werden beide Kameras frontal vor dem Fluglotsen schräg von oben leicht horizontal versetzt zueinander positioniert.

Die Anschaffung einer Grafikworkstation hat zudem die parallele Skeletterstellung zweier Videoströme ermöglicht. In Anbetracht dieser Änderung wurde die Skeletterfassung mit einer Frontalkamera durch die simultane Erfassung mit zwei horizontal versetzten Kameras ersetzt. Informationen über Position und Orientierung der Kameras ermöglichen die Fusionierung der 2D Skelette zu 3D Skeletten, sofern sämtliche Gelenke in beiden Aufnahmen nicht verdeckt sind. Daraus lässt sich die Position und Orientierung der 3D Skelette gegenüber den Kameras ableiten. Dies kann in die Position und Orientierung der aufgezeichneten Personen gegeneinander übertragen werden. Als Ergebnis liegen die Skelettdaten der Fluglotsen in 3D Form vor. Zusätzlich werden Informationen zu Positionen von Nase, Augen und Ohrmuscheln ausgegeben, welche als Ausgangspunkt für die Extraktion der Blickrichtung genutzt wurden.

Extraktion der Blickrichtung

Die Skeletterstellung liefert 3D Informationen über die Positionen von Nase, Ohren und Augen zu jedem Zeitpunkt der Aufnahme. Die Kenntnis über die relative Ausrichtung der Kamera zum Arbeitsplatz lässt damit Schlussfolgerungen über die Position und die Ausrichtung des Kopfes gegenüber dem Arbeitsplatz zu. Die Besonderheit der Lotsenarbeitsplätze zeichnet sich durch einen ständigen Wechsel der Blickrichtung zwischen dem frontal befestigten Radarschirm und dem auf der Arbeitsplatte befestigten Flugstreifendisplay aus. Die damit verbundene Abwendung des Kopfes von den oberhalb des Radarschirmes befestigen Kameras verhindert eine dauerhafte Pupillenerkennung. Allerdings kann die Blickrichtung abgeschätzt werden, ohne die Pupille selbst erkennen zu müssen. Durch die spezielle Arbeitsplatzgestaltung kann der mögliche Bereich des visuellen Fokus auf die Bildschirme, den zweiten Lotsen der Dyade oder einen Nachbarsektor eingegrenzt werden. Da die Skelette beider Teammitglieder in dieser Form vorliegen, sind Schlussfolgerungen zur Interaktion ableitbar.

Analyse der Bewegungsdaten und Extraktion der entsprechenden mathematischen Merkmale

Nach der Befragung von Fluglotsen der Centerniederlassung in Langen sowie der fachkundigen Beobachtung durch die im Projekt beteiligten Psychologen wurde ein Satz von relevanten arbeitsbezogenen Posen definiert, welcher die Beschreibung der alltäglichen Arbeitshaltung der Fluglotsen ermöglicht.

Zusätzlich wurden gemäß der relevanten Fachliteratur [48][101] körpersprachliche Signale untersucht, welche auf erhöhte Stresssituationen sowie psychische Belastung hinweisen.

Die Beobachtung der gleichen Person in den beiden unterschiedlichen Positionen (Executive und Planner) lässt erkennen, dass die Aufgaben in den beiden Positionen verschiedene Grundhaltungen nach sich ziehen. Da der Funkverkehr während der Simulation über ein in der Hand gehaltenes Mikrofon stattfindet, ist ein Aufstützen des Mikrofon-Armes auf den Tisch und damit verbunden ein häufigeres Vorlehnen der Lotsen zu beobachten. Eine grobe Analyse der Kopforientierung ergibt zudem unterschiedliche Foki der Aufmerksamkeit. Der Planner blickt deutlich öfter und länger auf den Bildschirm der Flugstreifen. Der Executive blickt die überwiegende Zeit auf den Radarschirm – unterbrochen von kurzen Eingaben auf den Flugstreifen. Beim Planner ist zudem in Phasen geringen Verkehrsaufkommens häufiger eine entspannte zurückgelehnte Haltung zu beobachten, beim Executive eher eine konzentrierte nach vorn gebeugte Haltung. Beobachtungen im Center in Langen und in München deuten darauf hin, dass diese vor allem der Simulation geschuldet ist, da dauerhaft eine mittlere bis hohe Auslastung der Fluglotsen provoziert werden sollte. Im realen Betrieb führen Phasen mit geringen Verkehrsaufkommen auch beim Executive zu einer entspannten, zurückgelehnten Haltung.

Zusätzlich zur Festlegung von Grundposen wurden in diesem Arbeitspaket Signale der Körpersprache analysiert, welche auf Stress oder Langeweile hindeuten können. Als Beispiele können hier das Bedecken des Mundes mit der Hand beim Gähnen oder der Griff zum Hals in einer Stresssituation genannt werden.

Der Vorteil der körpersprachlichen Signale besteht darin, dass sie unbewusst erfolgen und damit in den meisten Fällen von der ausführenden Person selbst nicht wahrgenommen werden. Eine Diskussion mit den Fluglotsen über dieses Thema brachte allerdings die Erkenntnis, dass es den Kollegen sehr wohl möglich ist, anhand der Körpersprache eines Lotsen festzustellen, ob dieser sich in einem erhöhten Stresszustand befindet und entsprechend darauf zu reagieren.

Die Auswertung eines Experiments mit Studenten in einem an die Fluglotsentätigkeit angelehnten Szenario (Experiment 2, siehe 1.5) zeigte einen Zusammenhang zwischen dem aktuellen Flugverkehr und körpersprachlichen Signalen auf. Ein bis zwei Sekunden vor der Erteilung eines Lotsenkommandos waren bei mehreren Versuchsteilnehmern Berührungen von Kinn, Mund oder Wange zu beobachten, was ein Indikator für erhöhte Konzentration sein könnte. Ein bis zwei Sekunden nach einer Kollision zweier Flugzeuge wurden einer oder beide Arme gehoben und die Hand zum Kopf geführt. Dieses Verhalten könnte damit auf Stress hindeuten.

Ein Nachteil der körpersprachlichen Analyse besteht jedoch darin, dass sie nur kontextabhängig erfolgen kann, da der psychologische Hintergrund nicht anhand der Videoaufzeichnungen identifiziert werden kann. Eine Berührung des Mundes als eventuelles Zeichen erhöhter Konzentration kann aufgrund eines hohen Verkehrsaufkommens stattfinden, ist aber ebenso bei Unterlast möglich, wenn über nicht zur Arbeit gehörenden Dinge nachgedacht wird. Die Wiederholung des studentischen Experiments mit Fluglotsen im Center in München zeigte keine vergleichbaren Ergebnisse. Weder vor Erteilung eines

Kommandos noch nach einem Zusammenprall zweier Flugzeuge erfolgten Änderungen der Körperhaltung. Ebenso ist bei der Auswertung der Simulationen aus Langen nicht gelungen, für alle Lotsen gültige Korrelationen zwischen Arbeitssituation und spezifischen körpersprachlichen Signalen zu ermitteln. Die Auswertung der Körpersprache als Indikator für Stress hat sich in diesem Fall als ungeeignet erwiesen.

Basierend auf dem ermittelten Satz an Basisposen wurden Verfahren der Datenanalyse und -clustering verwendet, um die während der Simulation aufgezeichneten Daten auf diese Posen hin zu untersuchen. Die Zusammenfassung von Posen erfolgt dabei auf Basis der Gelenkwinkel der 2D-Skeletts. Für je zwei Posen wurde eine Distanzfunktion als Summe der Differenzen der Gelenkwinkel der Posen definiert. Aufbauend auf dieser Distanzfunktion ist die Anwendung von Clusterverfahren möglich.

Als einfach zu implementierender und erfolgversprechender Algorithmus hat sich k-Medoids hervorgetan. Ausgehend von einem Satz von definierten Startposen wird eine Menge von Cluster erzeugt, welche durch je eine Pose als Repräsentanten dargestellt werden können. Jeder dieser Repräsentanten lässt sich durch spezifische Werte aller Gelenkwinkel beschreiben.

Es wurden zwei Ansätze für die Zerlegung der Menge aller Posen in Teilmengen getestet. Im ersten Ansatz wurde eine durch Beobachtung der aufgezeichneten Simulationen ermittelte Menge an häufig auftretenden sowie seltener – jedoch interessanter – Posen festgelegt. Anhand dieser Ausgangsposen erfolgte eine Zerlegung der kompletten Aufzeichnung einer Simulationssitzung je Fluglotse.

Der zweite Ansatz besteht in einer komplett automatisierten Zerlegung, welche die Menge der Posen systematisch in kleinere Cluster zerlegt, bis die Standardabweichung der Distanzen innerhalb der Cluster einen zuvor festgelegten Grenzwert nicht mehr überschreitet.

Betrachtet man die Mehrheit der Posen, so erzeugen beide Verfahren vergleichbare Ergebnisse. Tests mit verschiedenen Grenzwerten für die Abweichungen innerhalb der Cluster zeigen, dass es eine Untermenge von Posen gibt, welche übliche Arbeitshaltungen der Centerlotsen wiedergeben. Diese Grundhaltungen umfassen zwar nur etwa 13 Prozent der Cluster, enthalten jedoch bereits 90 Prozent der Frames der Aufnahme. Dieser statistische Zusammenhang war sowohl bei den Simulationen in Langen als auch bei dem studentischen Experiment zu beobachten. Die Wiederholung des studentischen Experiments mit Lotsen im Center wich jedoch hiervon ab. Die nachfolgenden Werte ergeben sich aus der Clusterbildung mit fünf Grad Gelenkwinkelabweichung als Grenzwert. Das studentische Experiment von vier Durchläufen zu je sieben Minuten Laufzeit zeigt zwischen 1 und 83 verschiedenen Posen (Durchschnitt 18, Median: 17), wobei zwischen zwei und vier Grundposen vorkamen. Die durchschnittliche Gesamtanzahl der Posen sank dabei mit jedem Wiederholungsdurchlauf von 29 Posen im ersten auf 18 Posen im vierten Durchlauf. Die Anzahl der Studenten, die den kompletten Versuchszeitraum über lediglich eine Pose eingenommen haben, stieg von anfangs 15 Prozent auf ein Wert von 34 Prozent im vierten Durchlauf. Das Experiment legt nahe, dass mit zunehmender Beschäftigung mit der Fluglotsenaufgabe die Anzahl der körpersprachlichen Reaktionen darauf sinkt. Dieser Zusammenhang könnte eine Erklärung für das Ergebnis der Wiederholung des Experiments mit Centerlotsen Im Center München sein. Hierbei haben 65 Prozent der Lotsen über den kompletten Versuchszeitraum lediglich eine Pose eingenommen. (Durchschnitt: 6, Median: 1). Bei den einstündigen Simulationen der Forschungsabteilung in Langen traten zwischen 143 und 206 verschiedene Posen auf (Durchschnitt: 174, Median: 169), wobei durchschnittlich 24 Posen 90 Prozent der Zeit abgedeckt haben. Diese Grundhaltungen unterscheiden sich im Wesentlichen durch die nach vorn oder hinten gerichtete Sitzhaltung sowie die Stellung der Arme zum Torso und zueinander. Ein Großteil der selten vorkommenden Posen

ist der Interaktion mit den Simulationsleitern geschuldet, nicht der Lotsentätigkeit selbst. Die bei dem studentischen Experiment erkannten Posen einer Denk- und Fehlerhaltung waren bei den Fluglotsen weder in der Simulation noch in der Wiederholung des studentischen Experiments zu beobachten.

Die Beobachtungen der Fluglotsen in Simulationsszenarien und im realen Centerbetrieb in Langen und München haben keinerlei Informationen über häufig auftretende Gesten geliefert, welche dem Arbeitskontext zuzuordnen wären. Die automatisierte Untersuchung der aufgenommen Daten in Bezug auf Gesten setzt die Hinzunahme der Fingergelenke in das 3D Skelett voraus. Hierfür kann das Verfahren von [147] zum Einsatz kommen, welches in OpenPose integriert ist. Die Fusion der Fingerdaten zu 3D Skeletten der Hand ist entsprechend des Verfahrens möglich, welches für das Gesamtskelett zum Einsatz kommt. Allerdings führt dies etwa zu einer Verdreifachung des Zeitaufwandes der Skelettierstellung bei Aufzeichnung einzelner Personen. Eine weitere Erhöhung des Zeitaufwandes ergab sich durch weitere an der Simulation beteiligte Personen, da jede im Bild erkannte Hand einen separaten Rechenschritt zur Erstellung der Finger nötig macht. Die Praxistauglichkeit einer automatisierten Gestenauswertung ist damit nach aktuellem Stand der Technik nicht gegeben.

1.2 Mimikanalyse (AP 4, AP 10, AP 14)

Extraktion von Mimik mittels Active Appearance Modellen

Um Gesichtsausdrücke zu erkennen, gibt es drei wichtige Schritte: (1) Gesichtsmerkmalextraktion, bei der hervorspringende Punkte auf einem Gesicht (z. B. Ecken des Mundes oder der Augen) automatisch aus unbekannten Bildern extrahiert werden; (2) FACS-Kodierung [34], d.h. Kodierung der einzelnen Muskelbewegungen (Action Units, AU) im Gesicht und (3) Emotionserkennung/Repräsentation auf der Grundlage der AUs allein oder in Kombination mit den Landmarks und deren Umgebung.

Das untersuchte Verfahren zur Gesichtsmerkmalextraktion war eine Variante der Active Appearance Model Methode (AAM, [90]), die einen inversen Kompositionsalgorithmus verwendet, um fortschreitend Landmarks an das Gesicht anzupassen, indem der Rekonstruktionsfehler zwischen dem generierten Gesicht des Modells und einem mittleren Gesicht, das aus einem Satz von Trainingsbildern extrahiert wurde, minimiert wird. Der Hauptvorteil des inversen Kompositionsalgorithmus gegenüber dem klassischen Lucas-Kanade-Algorithmus besteht darin, dass die meisten Informationen über die Transformation vorverarbeitet werden können, so dass nur eine minimale Rechenlast während des Anpassungsprozesses übrig bleibt und Echtzeitberechnungen möglich sind.

Der Hauptnachteil ist, dass die AAMs in ihrer ursprünglichen Formulierung empfindlich gegenüber verschiedenen Transformationen sind, wie etwa Beleuchtungsveränderungen, Verdeckungen oder vor allem in Bezug auf die Generalisierung bzgl. verschiedener Gesichter. Der Algorithmus wird auf bestimmten Gesichtern trainiert und generalisiert nur schlecht auf neue Gesichter. Im Rahmen dieses Arbeitspakets haben wir verschiedene Methoden untersucht und umgesetzt, die es ermöglichen, partielle Verdeckungen zu behandeln: Projekt-Out Inverse Compositional (POIC), Simultaneous Inverse Compositional (SIC), Efficient Robust Normalization (ERN) und Robust Simultaneous Inverse Compositional (RSIC). Experimente auf der Radboud-Gesichtsdatenbank mit künstlich erzeugten Okklusionen führten zu der Schlussfolgerung, dass die ERN-Methode den besten Kompromiss zwischen Robustheit gegenüber Okklusion und Rechenkomplexität bietet, gemessen an dem durchschnittlichen Fehler pro Pixel

zwischen den tatsächlichen unverdeckten Bildern und der Prognose des Modells von den verdeckten Bildern. Daher wurde ERN zur standard fitting Methode gewählt.

Hinsichtlich der Robustheit gegenüber Beleuchtungsänderungen untersuchten wir mehrere Vorverarbeitungsverfahren: Tan-Triggs-Vorverarbeitung, Multiscale Retinex Filter, Histogramm-Matching und Distance Maps. Untersuchungen an der erweiterten Yale-B Datenbank, die Gesichter unter verschiedenen Lichtbedingungen enthielten, zeigten, dass die Distance Maps in Kombination mit ERN die beste Leistung bei schlechten Lichtbedingungen mit einer Konvergenzrate von 200 Epochen bis zu 80% aufwiesen. Diese Kombination wurde zur Standardverarbeitungspipeline in unserer Implementierung. Wir haben auch untersucht, wie 2.5D-Modelle (die implizit die 3D-Struktur eines Gesichts berücksichtigen, ohne tatsächlich die Transformation in diesem Raum zu berechnen) die Leistung des AAM-Ansatzes verbessern könnten, allerdings konnte keine Verbesserung erzielt werden.

Das für die Ausbildung der AAM gewählte Wahrnehmungsmuster ist die Candide-3-parametrisierte Gesichtsmaske. Diese Maske wird in der MPEG-4 Gesichtsanimationsparameter-Norm verwendet und wird standardmäßig für die Gesichtserkennung in Kinect-Geräten verwendet. Ihr Hauptinteresse ist, dass die Landmarks dem AU-Vektor-Endpunkt des FACS-Codierungssystems entsprechen, was die Extraktion von AU-Vektoren aus der Verschiebung dieser Punkte erleichtern sollte. Allerdings erwies sich dieser Ansatz als schwierig, da die Schwellenwerte, die den 5 AU-Intensitäten entsprechen, schwer in einer generischen Weise einzustellen sind: unterschiedliche Gesichtsmorphologien führen zu unterschiedlichen Verschiebungsmustern, und ein einzigartiger Satz von Regeln konnte nicht gefunden werden.

Entwicklung eines Deep-Learning Active Appearance Modells zur Extraktion von Mimik

Um das beschriebene Problem zu überwinden, verwendeten wir Techniken aus dem Maschinellen Lernen, um die Zuordnung zwischen Landmark-Verschiebungen und AU-Aktivierung robuster schätzen zu können. Unsere Experimente zeigten, dass eine bessere Vorhersagegenauigkeit erreicht werden kann, wenn ein feedforward neuronales Netzwerk verwendet wird, welches die Verschiebung der Landmarks mit der entsprechenden AU-Aktivierung auf der RadBoud-Gesichtsdatenbank abzubilden gelernt hat, obwohl es an Generalisierung fehlte, da nicht genügend Daten verfügbar waren.

Diese Vorgehensweise wurde durch ein tiefes Convolutional Neural Network (4 Faltungsschichten mit ReLU-Transferfunktion und Dropout, jeweils gefolgt von Max-Pooling, einer vollständig verbundenen Schicht und einer Softmax-Schicht für die Vorhersage) direkt auf die Rohbilddaten ausgedehnt, um die entsprechenden grundlegenden Emotionen (Anger, Disgust, Fear, Happiness, Sadness, Surprise, Neutral) vorhersagen zu können. Zu diesem Zweck sammelten wir so viele Gesichtsausdruck-Datensätze wie möglich (CCK+, MMI, UNBC-Pain, CFEE, DISFA, Kaggle und IMFDB), die eine riesige Vielfalt von Ansichten in kontrollierter Pose oder in natürlicher Umgebung zeigen, und erhielten so insgesamt 180.000 Bilder. Das tiefe Netzwerk wurde trainiert, um die grundlegende Emotion in diesen Bildern vorherzusagen. Wir erzielten eine Genauigkeit in der Nähe des Standes der Technik auf jedem Datensatz (z.B. 90,2% gegenüber 96,4% bei CCK+; 68,4% gegenüber 69,4% bei Kaggle; 91,16% gegenüber 93,8% bei DISFA). Diese Ergebnisse deuten darauf hin, dass ein End-to-End-Ansatz für Gesichtsausdrucks-Erkennung ausreichend gut generalisiert.

Der Nachteil dieses Ansatzes ist, dass das tiefe Netzwerk auf rohen Bildern (einschließlich irrelevanter Regionen wie dem Haar oder dem Hintergrund) trainiert wird, was viel unnötiges Training und Testzeit

verbraucht. Wir haben daher untersucht, ob die Ausgabe des AAM verwendet werden kann, um relevante Gesichtsmerkmale zu isolieren und so nur die relevanten Bereiche in die tiefen Netzwerke einzubringen und daher ihre Geschwindigkeit und Verallgemeinerungsgenauigkeit zu verbessern. Leider hat diese Lösung zu keiner Verbesserung geführt, meistens wegen der Ungenauigkeit des AAM. Wir entschieden deshalb, uns auf die Verwendung von tiefen neuronalen Netzwerken zu konzentrieren, um direkt aus den Beispielen die Abbildung zwischen Gesichtsbildern und den emotionalen Zustand zu lernen. Das Ergebnis dieses AP ist ein trainierter Klassifikator, der emotionale Zustände auf unbekannten Bildern mit einer Genauigkeit in der Nähe des Standes der Technik auf begrenzten Datensätzen produziert, deren Robustheit für Nicht-Standard-Gesichter aber verbessert werden könnte.

FACS Codierung von Gesichtsausdrücken

Die bisher benutzten Datensätze (CCK+, MMI, UNBC-Pain, CFEE, DISFA, Kaggle und IMFDB) sind nicht ausreichend, um eine befriedigende Robustheit des Klassifikators in Bezug auf Beleuchtung, Pose, Geschlecht und andere natürliche Variationen bei der Interaktion mit natürlichen Bildern zu gewährleisten. Deshalb haben wir zusätzliche Datensätze (SEMAINE und BP4D-spontaneous) mit einer hohen Anzahl von annotierten Gesichtern hinzugefügt (220.000 Gesichter in BP4D). Durch die zusätzlichen Daten konnten wir tiefere Netzwerke nutzen und komplexere und robuste Darstellungen lernen. Zusätzlich ist BP4D mit FACS annotiert (wie auch CCK+, MMI, PAIN und DISFA).

In FACS ist die Variation einer bestimmten Gesichtsregion (z.B. Ecken des Mundes oder der Augen) als Action Unit (AU) beschrieben, welche die Abweichung von einem neutralen Ausdruck beschreibt. Obwohl es in der FACS-Theorie 69 AUs gibt, sind 28 von ihnen für die Emotionserkennung hilfreich. Das EMFACS (Emotion FACS) System [36] erlaubt es, AU-Aktivierungen auf Grundemotionen abzubilden: zum Beispiel ist Happiness durch die gleichzeitige Aktivierung von AUs 6 (Wangenerhöhung) und 12 (Lippenwinkelzieher) gekennzeichnet. Ein generelles Problem beim Lernen besteht aber darin, dass die verschiedenen Datensätze nicht nur eine unterschiedliche Anzahl von Bildern enthalten, sondern auch verschiedene annotierte AUs haben: CCK+ beschreibt 30 AUs auf nur 593 Bildern und BP4D enthält 220.000 annotierte Bilder, aber nur mit 12 AUs: 1, 2, 4, 6, 7, 10, 12, 14, 15, 17, 23, 24.

Unser ursprünglicher Ansatz bestand darin, Ensembles von neuronalen Netzen zu verwenden, wo verschiedene Klassifikatoren auf verschiedene Datensätze trainiert werden und die endgültige Entscheidung mit einer Mehrheitsentscheidung erfolgt. Die Hauptschwierigkeit bestand somit sicherzustellen, dass die Klassifikatoren unabhängig sind. Leider war dieser Ansatz nicht erfolgreich, da die Datensätze zu unausgewogen sind: DISFA (30%), BP4D (42%) und SEMAINE (25%) haben mit Abstand die meisten Trainingsbeispiele, was die auf CCK +, MMI und PAIN trainierten Netzwerke in der Praxis nutzlos macht. Außerdem ist SEMAINE eine Untermenge von BP4D (obwohl mit verschiedenen AUs), so dass SEMAINE am Ende überflüssig war. Als günstigste Option erschien uns die Aggregierung von BP4D und DISFA zu einem kombinierten Datensatz und um lediglich ein einzelnes tiefes neurales Netzwerk darauf zu trainieren, da sich die annotierten AUs in DISFA und BP4D weitgehend überschneiden: Die Ausgabe des Netzwerks sollte das Auftreten der 12 AUs vorhersagen, die in BP4D vorhanden sind. Für die in DISFA nicht vorhandenen AUs (7, 10, 14, 23, 24) wurden sie als abwesend angesehen, was aber die Genauigkeit des Netzwerks beeinflussen könnte.

Neuronale Netzwerke benötigen Eingabebilder mit fester Größe. In Übereinstimmung mit anderen Ansätzen entschieden wir uns, den Haar-Kaskadenklassifizierer für die Gesichtsregion aus der dlib-Bibliothek (http://www.dlib.net) in allen Frames in unserem Datensatz anzuwenden und die Eingabe auf die

Größe von 96x96 Schwarz-Weiß zu normieren. Wir konnten 310.626 Gesichter aus den beiden Datensätzen extrahieren: unser kombinierter Datensatz umfasst daher 310.626 Bilder (Array der Größe (310.626, 96, 96, 1)) und Labels (Array der Größe (310.626, 12)). Die Validierungs- und Testsätze umfassten nur jeweils 5% der Gesamtproben, da das Ziel darin besteht, den bestmöglichen Klassifikator für das Macelot-Projekt zu erstellen. Leider ist der resultierende Datensatz bzgl. der AUs nicht ausgeglichen: Einige AUs kommen häufiger vor als andere. Abbildung 3 zeigt die Häufigkeitsverteilung der AUs im kombinierten Datensatz an. Dieses Klassenungleichgewicht kann einen signifikanten Einfluss auf die Leistung des Klassifikators bei weniger häufigen AUs haben. Die meiste Zeit in diesem AP wurde für die Suche nach der optimalen Struktur des neuronalen Netzes (Anzahl der Schichten, Meta-Parameter) eingesetzt. Die Siegerarchitektur wird weiter unten vorgestellt.

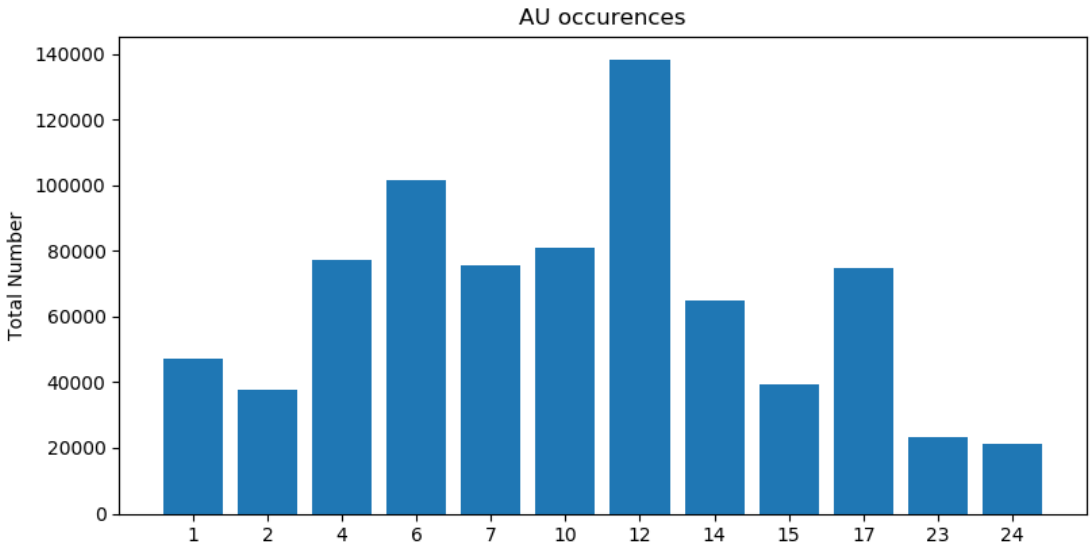


Abbildung 3: Anzahl der annotierten Bilder im kombinierten Datensatz (DISFA und BP4D) für jede AU.

GPU basiertes Modell zur Extraktion von Mimik-Merkmalen

Da wir von Active Appearance Models zu einem reinen Deep-Learning-Ansatz übergegangen sind, konnten wir während des Projekts von den enormen Fortschritten bei Deep-Learning-Frameworks profitieren: Theano, Torch, Tensorflow und andere Verfahren sind inzwischen extrem leistungsfähig hinsichtlich der Simulation großer neuronaler Netzwerke auf NVIDIA Karten. Die Verwendung der Tensorflow-Bibliothek [1] ermöglichte es uns, die Trainingszeit des neuronalen Netzwerks im Vergleich zur CPU-Variante um den Faktor 50 zu reduzieren. Dies ermöglichte uns, das Netzwerk erfolgreich auf dem riesigen kombinierten Datensatz zu trainieren. Das Training benötigte dennoch drei Tage Rechenzeit.

Optimiertes Modell zur FACS Codierung von Mimik-Merkmalen

Das ausgewählte neuronale Netzwerk (dargestellt in Abbildung 4) besteht aus 4 Faltungsblöcken, die jeweils aus 2 Faltungsschichten (Kerngröße 3x3, ReLU-Aktivierungsfunktion) und einer Max-Pooling-Schicht (2x2) bestehen. Eine Dropout-Schicht mit wird nach dem Max-Pooling hinzugefügt. Nach 4 solcher Faltungsblöcke mit zunehmender Anzahl von Merkmalen (32, 64, 126 und 256) wird der letzte Tensor (6x6x256) zu einem Vektor von 9.216 Elementen abgeflacht und auf eine vollständig verbundene Schicht von 500 Neuronen projiziert. Die Ausgabeschicht hat 12 Neuronen, die die Sigmoid-Akti-

vierungsfunktion verwenden, wobei jede eine der 12 AU repräsentiert, die in dem kombinierten Datensatz vorhanden sind. Das Netzwerk hat insgesamt 5.786.192 trainierbare Parameter (Gewichte und Verzerrungen). Die verfügbaren GPUs unserer Rechner können mit so einem mittelgroßen tiefen Netzwerk umgehen.

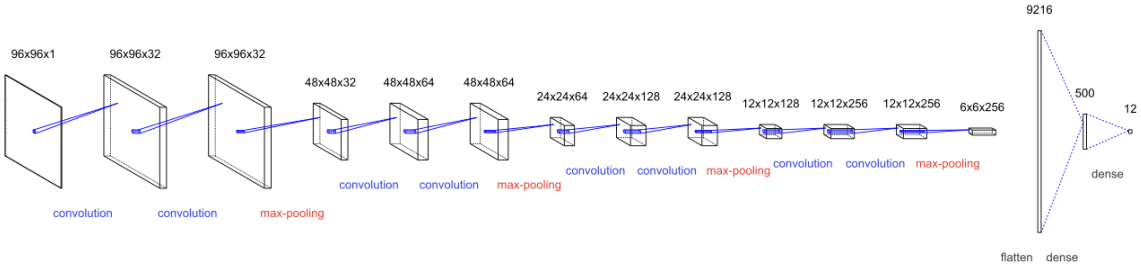


Abbildung 4: Beschreibung des deep convolutional neuronalen Netzwerks, das zur Vorhersage der AU-Aktivierung verwendet wird. Zwei Faltungsschichten werden angewendet, bevor eine Max-Pooling-Schicht die räumlichen Dimensionen reduziert. Dieser Block aus drei Schichten wird viermal wiederholt. Die letzte Faltungsschicht wird dann abgeflacht und auf eine vollständig verbundene Schicht gekoppelt und schließlich auf die Sigmoid-Ausgabeschicht projiziert. Dropout-Layer sind der Einfachheit halber weggelassen.

Da das Lernproblem eine Multi-Label-Klassifizierung ist (mehrere AUs (Action Units) können in einem einzigen Bild vorhanden sein), verwenden wir die Sigmoid-Aktivierungsfunktion für die Ausgabeschicht (von 0.0 für bestimmte Abwesenheit bis 1.0 für bestimmte Präsenz, alles dazwischen ist unsicher) und binär Kreuz-Entropie für die zu minimierende Verlustfunktion (die wahre Markierung ist 0 für AU Abwesend, 1 für AU Vorhanden). Der Ausgabe-Score für jedes AU stellt außerdem die Wahrscheinlichkeit dar, dass eine AU detektiert wird, d.h. von der Normalposition abweicht: zur Klassifikation haben wir ein Schwellenwert von 0,5 auf diesen Score gesetzt (d.h. eine AU wird detektiert, wenn der entsprechende Score über 0,5 liegt). Das Modell wurde über 120 Epochen mit Stochastic Gradient Descent (SGD) auf Minibatches von 128 Proben mit einer Lernrate von 0,01 und einem Nesterov-Momentum von 0,9 trainiert. Abbildung 5 (links) zeigt an, dass das Netzwerk die Trainingsdaten erfolgreich gelernt hat (Testverlust 0,02) und nur sehr leicht übersteuert wurde. Die Fehlerrate für jede AU (Abbildung 5 rechts) ist sehr niedrig (von 1% bis 5%), aber das Klassenungleichgewicht hat einen leichten Effekt. Auf dem Test-Set berechneten wir auch den F1-Score für jede AU einzeln, der von 0,93 bis 0,99 reichte.

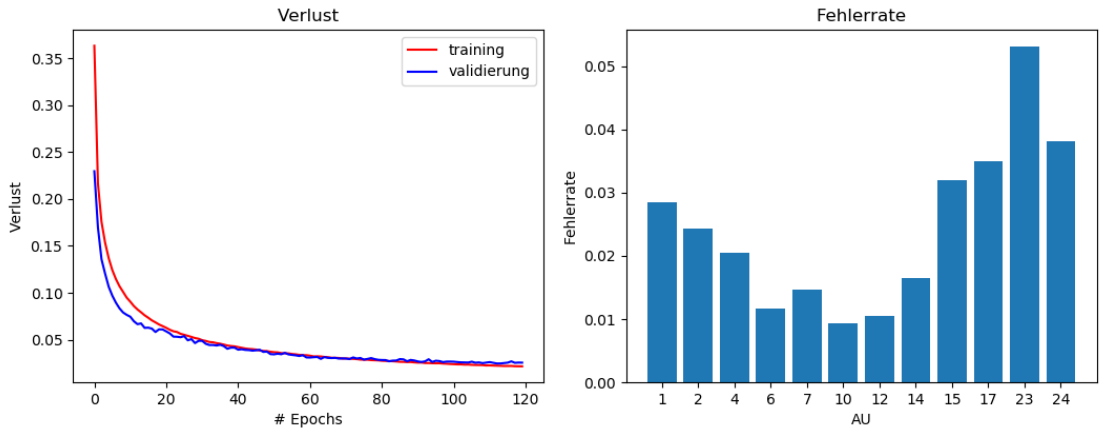


Abbildung 5: Links: Trainings- und Validierungsverluste in den 120 Trainingsepochen. Rechts: Fehlerrate auf dem Test-Set für jede AU

Sobald das FACS-Netzwerk erlernt ist, können Emotionen direkt unter Verwendung der EMFACS-Korrespondenzmatrix vorhergesagt werden: Happiness (6, 12), Sadness (1, 4, 15), Surprise (1, 2), Fear (1, 2, 4, 7), Anger (4, 7, 23), Disgust (15). Man kann sehen, dass Disgust nur auf AU 15 beruht, daher wäre seine Erkennung extrem empfindlich gegenüber falsch positiven (oder negativen) Ergebnissen. In ähnlicher Weise sind Überraschung und Angst durch die Anwesenheit von AU 4 und 7 unterscheidbar, während Angst und Wut von der korrekten Erkennung von AU 1, 2 oder 23 abhängen. Um die Übereinstimmung zwischen einer prototypischen Emotion (z. B. Traurigkeit 1, 4, 15) und ein erkanntes Muster von AUs (zum Beispiel 1, 4, 12) zu messen, berechnen wir ihren F1-Wert. Der F1-Wert ist das harmonische Mittel der Genauigkeit (Anzahl der wahren positiven Werte (1, 4) dividiert durch die Gesamtzahl der Elemente, die vom Klassifikator (1, 4, 12) als positiv klassifiziert werden) und Recall (Anzahl der wahren positiven (1, 4) dividiert durch die Gesamtzahl der Elemente, die tatsächlich zur positiven Klasse gehören (1, 2, 15)). Wenn der F1-Wert 1 ist, werden nur die prototypischen AUs der Emotion erkannt (1, 4, 15). Jede verpasste AU (15) oder zusätzliche AU (12) verringert den F1-Wert. Der Wert 0,5 entspricht dem Zufallslevel.

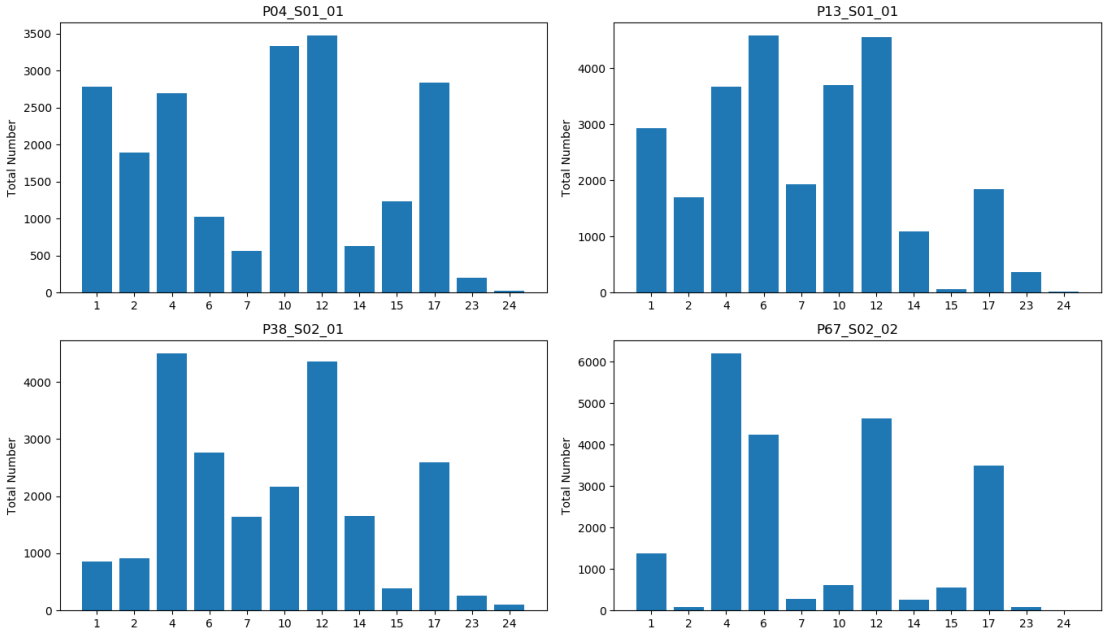


Abbildung 6: Gesamtzahl der vorhergesagten AUs in allen Frames der 4 Demo-Videos.

Zusätzlich zu dem Videomaterial der Fluglotsen wurde an der TU Chemnitz ein Experiment zur simulierten Fluglotsensituation durchgeführt, um den Einfluss der Emotionen auf die Leistung der Aufgabe zu untersuchen. Dafür wurde eine Emotionsinduktion durchgeführt in der 70 Teilnehmer emotionsgeladene Videos sahen. In jeder Sitzung waren 7 Minuten besonders kritisch für die Emotionserkennung und die entsprechenden Frames wurden extrahiert (10.500 Frames bei 25 fps). Zum Zweck der Demonstration konzentrieren wir uns hier auf 4 spezielle Sitzungen: P04 S01, P13 S01, P38 S02 und P67 S02. Abbildung 6 zeigt die Gesamtzahl der AU-Erkennungen während der vier Demo-Videos. Die Verteilung der erkannten AUs ist in diesen “freien” Videos im Vergleich zu den Trainingsdaten unterschiedlich. Insbesondere werden AU 23 und 24 fast nie detektiert, während sie häufig in den Trainingsdaten entdeckt wurden. Bestimmte AUs wie 4 (Brow Lowerer) und 12 (Lip Corner Puller) werden dagegen sehr oft in den Videos gefunden. AU 4 kann durch die von den Probanden getragene Brille gestört sein, was die korrekte Erkennung von Augenbrauenbewegungen beeinträchtigt. AU 12 kann auch durch die

Tatsache beeinflusst werden, dass die Versuchspersonen die Möglichkeit haben, während der Experimente zu sprechen, so dass sich der Mund häufig öffnet und schließt, die Lippenecken sich bewegen, und die Sicht der Person nicht frontal ist. Für das Netzwerk, das nur frontale Gesichter gesehen hat, kann die schräge Blickrichtung dazu führen, dass die AU für Lippenecken anspricht.

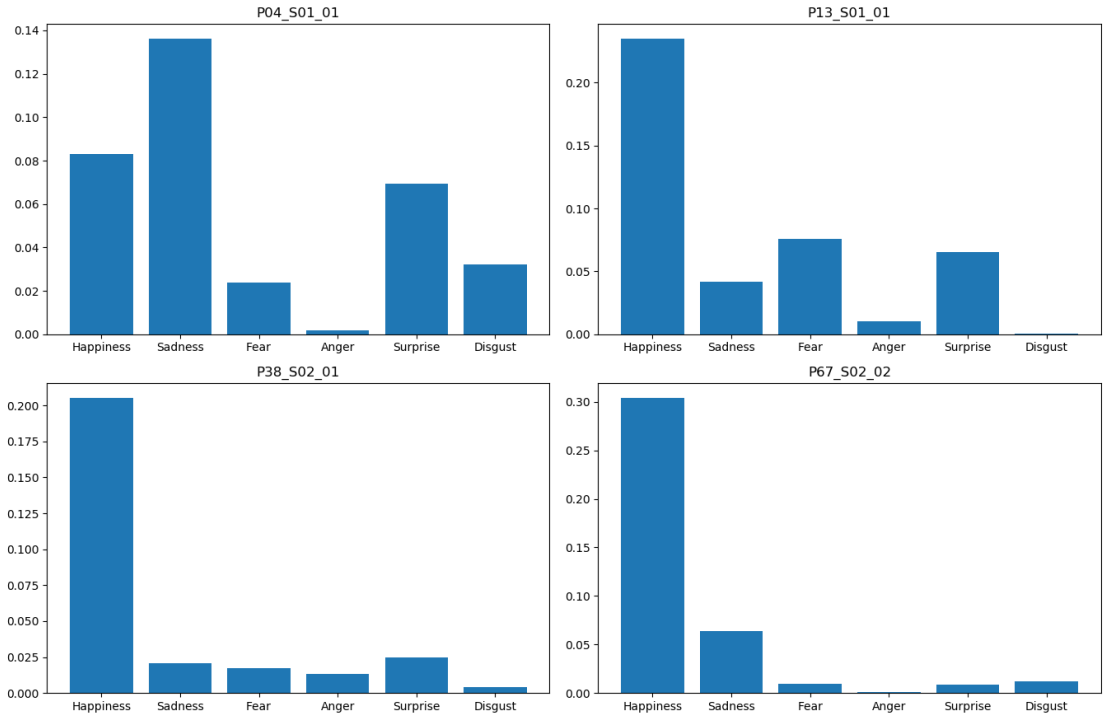


Abbildung 7: Vorhersage der Emotion in den Bildern der 4 Demo-Videos.

Abbildung 7 zeigt die Verteilung von Frames, die positiv mit einer bestimmten Emotion verbunden sind, unter Verwendung der EMFACS-Korrespondenzmatrix. Da die Testvideos ziemlich lang sind (7 Minuten), werden mehrere Emotionen erkannt, aber Glück und Traurigkeit werden häufiger vorhergesagt, vermutlich, da dies die induzierten Emotionen in diesen Sitzungen waren. Die Vorhersagen müssen allerdings mit Vorsicht behandelt werden: viele Frames sind korrekt detektiert (Siehe Abbildung 8) aber es gibt immer noch false positives. Während unser tiefes neuronales Netzwerk in der Lage ist, AU-Aktivierung auf standardisierten Datensätzen genau zu erkennen, erweist sich seine Anwendung auf unbeschränkte Videoeinstellungen immer noch als schwierig, da das aufgenommene Videomaterial stärker von den Bedingungen der Trainingsmenge abweicht.

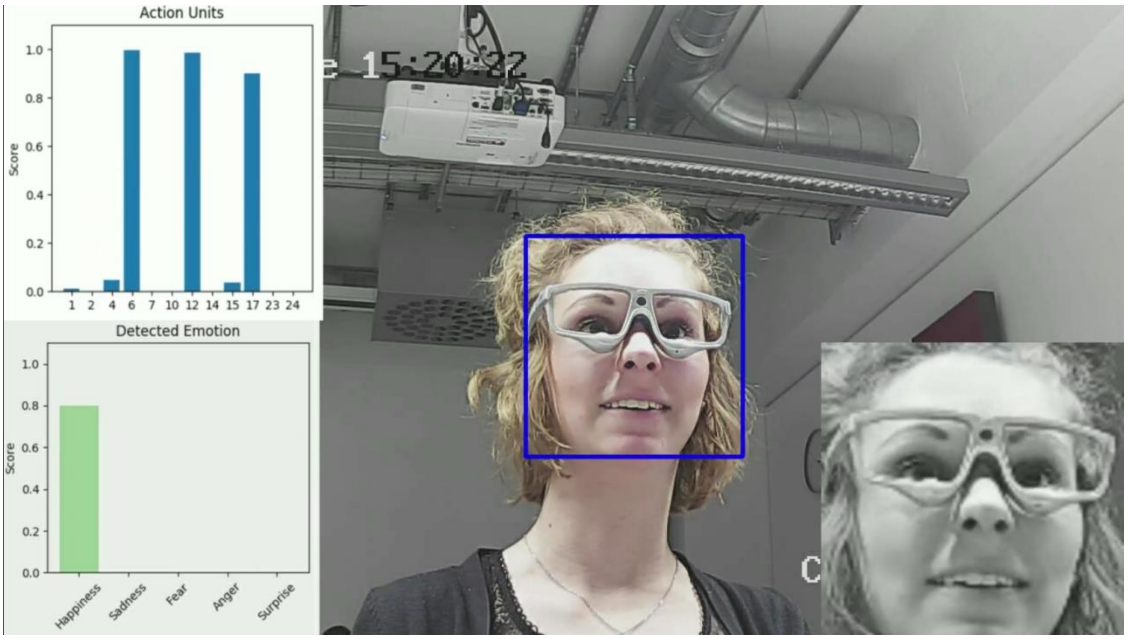


Abbildung 8: Beispiel für eine korrekte Erkennung der Emotion Glücklich für P04 S01. Die Gesichtserkennung ist durch das blaue Quadrat gekennzeichnet und die Eingabe in das neuronale Netzwerk befindet sich in der unteren rechten Ecke. Der Erkennungsstand der 12 AU.

Modell zur Detektion zur Erfassung der zeitlichen Entwicklung von Gesichtsausdrücken

Die Verwendung der EMFACS-Korrespondenzmatrix in den vorherigen Arbeitspaketen erzwingt eine statische Assoziation zwischen den FACS-Anmerkungen für alle Bilder und der zugrundeliegenden Emotion. Über längere Zeiträume kann die Emotion durch Mittelung und Glättung der Vorhersagen erkannt werden, wobei potentiell interessante dynamische Effekte vernachlässigt werden. Das Ziel dieses Arbeitspakets war es, zu untersuchen, ob die zeitliche Abfolge von detektierten AUs mehr Informationen als einzelne Frames trug. Wir modifizierten das vorherige CNN, um Repräsentationen auf hoher Ebene (entweder die 12 AUs der Ausgabeschicht oder die 500 Neuronen der letzten vollständig verbundenen Schicht) der Eingabebilder zu extrahieren und sie einem rekurrenten neuralen Netzwerk (LSTM) zuzuführen, dessen Aufgabe es ist die dominante Emotion in kurzen Videosequenzen zu klassifizieren (verfügbar im EmotioNet-Datensatz). Wir untersuchten verschiedene Varianten dieser Architektur (Anzahl der Neuronen, maximale Länge einer Sequenz, Hyperparameter), konnten jedoch keine signifikante Verbesserung der Genauigkeit bei der Emotionserkennung feststellen (im Vergleich zu einer einfachen Mehrheitsabstimmung über alle Frames anhand des vorherigen CNN). Es verbleibt unklar, ob dieser Mangel an Verbesserung auf der Methodik oder auf der Tatsache beruht, dass die Sequenzen von Bildern nicht mehr Informationen als einzelne Bilder tragen.

1.3 Audioanalyse (AP 5, AP 11, AP 15)

Konzeption des Evaluationskorpus und Erfassung von Audiodaten

Mit Beginn des Vorhabens wurde zunächst die Struktur eines Sprachdatenkorpus für das Fluglotsenszenario konzipiert und ein entsprechendes Audio-Repository angelegt. Die im Vorfeld festgelegten Typen von Attributen dienen dem Zweck modalitätsübergreifenden Untersuchungen bezüglich akustischer und linguistischer Besonderheiten:

- Sprecher ID: 1; …; n

- Geschlecht: male; female
- Rolle des Fluglotsen: executive; planner
- Einsatz von Sprache: radio; no_radio
- Sprache: deu; eng
- Mehrsprachigkeit: monolingual; bilingual

Mittels der beschafften Aufnahmetechnik wurden beim Kooperationspartner DFS (Deutsche Flugsicherung GmbH) in Langen Audiomitschnitte typischer sowie kritischer Situation in den Simulationen vorgenommen. Sprache konnte mittels eines Ansteckmikrofons aufgenommen werden, welches mit einem Sender-Empfänger System gekoppelt war (Sennheiser ew 112). Die Monosignale wurden mit einer Frequenz von 48 kHz bei 24 bit Auflösung durch ein mobiles Aufnahmegerät (Roland R-88) erfasst. Bei der Aufnahme mehrerer Fluglotsen innerhalb einer Simulation erhielt jeder Fluglotse einen eigenen Audiokanal mit entsprechendem Mikrofon; eine Simulation dauerte im Schnitt 60 Minuten. Die Audioaufnahmen wurden anschließend mit 16kHz und 16 bit in das WAV-Format transkodiert und in das Repository der Medieninformatik zur intellektuellen Annotation integriert.

Insgesamt fanden 12 Simulationen statt, wobei die Menge an Teilnehmern je Versuch variierte. Einige Fluglotsen nahmen an mehreren Simulationen teil (insgesamt 11).

Transkription für den Evaluationskorpus

Linguistische Untersuchungen sowie Diskursanalysen hinsichtlich des Fluglotsenbetriebes erfordern zunächst eine verschriftlichte Form von Äußerungen der akquirierten Audiomittschnitte aus den Simulationen. Für die effiziente manuelle Transkription kam das Werkzeug FOLKER [136] des IDS (Institut für Deutsche Sprache Mannheim) zum Einsatz (siehe Abbildung 9). Hierbei wurde die Transkriptionskonvention GAT 2 [144] berücksichtigt. Darüber hinaus erfolgte die intellektuelle Annotation von Sprecher und Sprache mittels festgelegter Metadatentypen. Somit können u.a. die Rolle des Fluglotsen und der Einsatz von Sprache beschrieben werden.

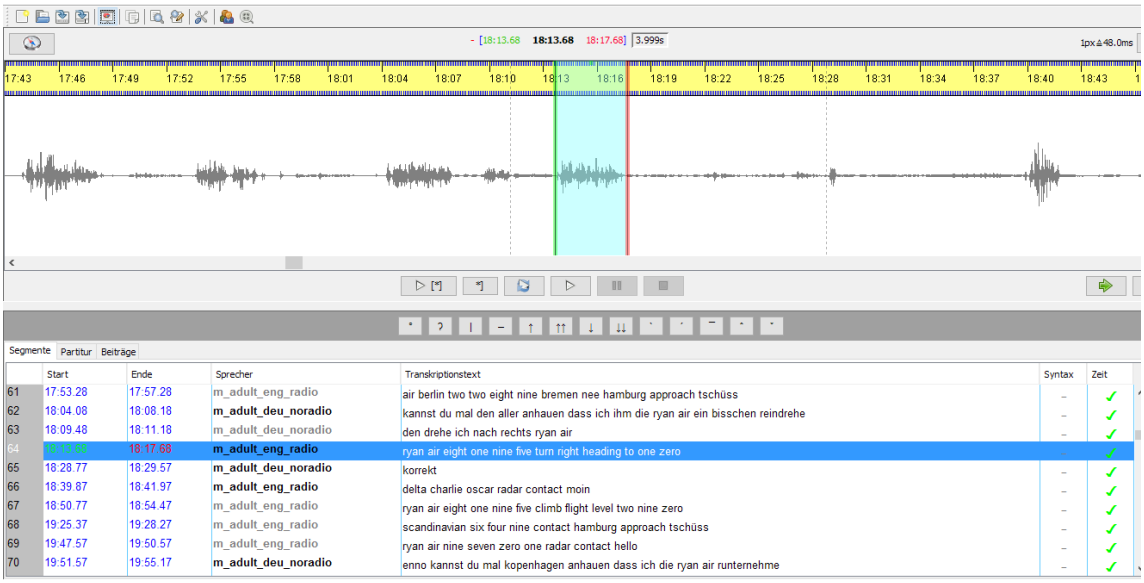


Abbildung 9: Beispiel einer Transkription von Äußerungen eines Fluglotsen (Executive Controller) mittels FOLKER

Tabelle 1 gibt einen allgemeinen Überblick der transkribierten Aufnahmen mit Unterscheidung hinsichtlich der Rolle des Fluglotsen. Darüber hinaus sind Kennzahlen für Fluglotsendyaden gegeben. Alle Fluglotsen werden hierbei durch Executive und Planner unterschieden. Im Fall von Dyaden sind alle Fluglotsen-Paare zusammengefasst; entsprechend ist die Zahl der Dyaden gleich der Planner. Die Tabelle verdeutlicht, dass zudem eine Reihe von Fluglotsen-Sprachdaten transkribiert wurden, welche nicht im Dyaden-Kontext stehen.

Beschreibung	**Fluglotsen**	**Executives**	**Planner**	**Dyaden**
# Simulationsaufnahmen	29	21	8	8
# Äußerungen	8.196	6.473	1.713	3.655
# Wörter	65,2k	53,8k	11,3k	30,5k
# Vokabular	3.166	2.481	1.674	2.350
Dauer gesamt [hh:mm]	05:43	04:44	00:59	02:34
Durchschnittliche Dauer pro Äußerung [s]	2,5	2,7	2,1	2,5

Tabelle 1: Datenbeschreibung transkribierter Sprache aus Fluglotsen-Simulation

Erstellung von Metadaten für den Evaluationskorpus

Um Sprachtechnologien auf verschiedenen Ebenen auswerten zu können, wurden im weiteren Verlauf des Vorhabens Metadaten erzeugt, welche sich insbesondere auf die Kommunikation von Fluglotsen beziehen. Dies schafft zudem eine weitere Grundlage zu modalitätsübergreifenden Untersuchungen hinsichtlich physiologischer Signale und Verhaltensmuster.

Durch die Erfassung der Fluglotsen hinsichtlich aufgabenbezogener Dialoge und deren Klassifizierung auf Dyaden-Ebene, könnten perspektivisch die Qualität der Zusammenarbeit, das Arbeitsklima oder die Belastung der Teams ausgewertet werden. Als Ansatz wurde hierzu das Annotationsschema SWD-DAMSL (Switchboard - Dialog Act Markup in Several Layers) [64] angewendet, um Dialoge im Arbeitsumfeld zu beschreiben. Äußerungen werden dabei auf vier Ebenen erfasst:

- *Communicative Status* beschreibt Ereignisse, welche von der Standardkommunikation abweichen. Dazu zählen *abandoned*, *uninterpretable*, *self-talk* und *third party talk*.
- *Information Level* beschreibt den Fokus des Dialoges und wird entsprechend annotiert mit *task*, *task-management*, *communication-management*, *other-level*
- *Forward-Looking-Function* beschreibt die Auswirkung auf den weiterführenden Dialog. Hierzu wurden 21 Tags verwendet, wie beispielsweise *yes/no question*, *action directive* (Anweisung) und *statement-opinion* (Meinungsäußerung)
- *Backward-Looking-Function* beschreibt Reaktionen auf vorherige Äußerungen. 24 Tags wurden hierzu aufgegriffen, u.a. *yes answer*, *no answer*, *agreement*, *reject* und *signal none understanding* (bspw. „Wie bitte?“)

Für eine zunächst manuelle Annotation der aufgenommenen Simulationsläufe kam die Software ELAN [161] des Max-Planck-Instituts für Psycholinguistik zum Einsatz. Für je einen Simulationslauf der Fluglotsen wurde das Transkript des entsprechenden Executives und Planners importiert und der Dialog mittels des Annotationsschemas SWD-DAMSL beschrieben. Als Beispiel sei die Datenauswertung des Dialoges auf Informationsebene vom 22. September 2015 in (Abbildung 10) illustriert. Hierbei ist die Relevanz der Dialogbeschreibung deutlich zu erkennen: die eigentliche Durchführung der Aufgabe (task) erfordert vergleichsweise eine hohe Anforderung der Aufgabenplanung und -regulierung (task-

management); ebenso spielt das Aufrechterhalten der Kommunikationsqualität (communication-management) eine wichtige Rolle. Der größte Anteil des Dialoges bezieht sich jedoch nicht auf die eigentliche Aufgabe (other).

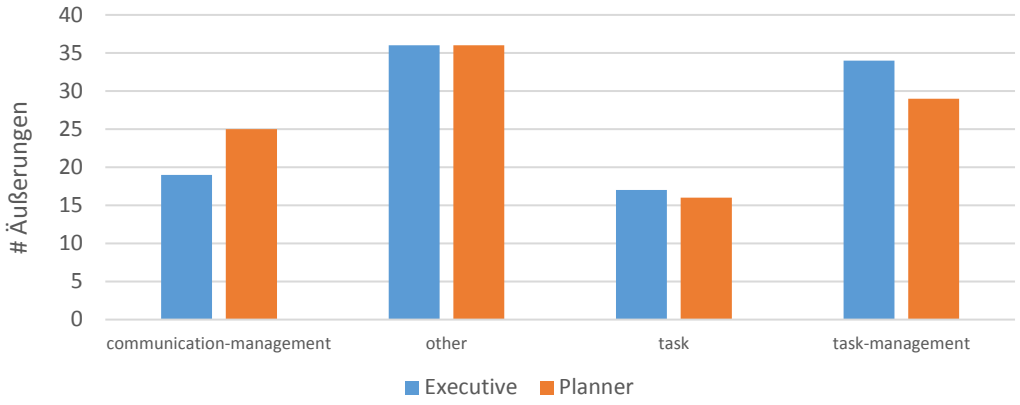


Abbildung 10: Datenauswertung des Dialogs während eines Simulationslaufes vom 22. September 2015 der Fluglotsen (Dyade) auf Informationsebene.

Die vollständig automatisierte Diskursanalyse auf verschiedenen Kommunikationsebenen innerhalb eines Dialogaktes erfordert sowohl eine automatische Spracherkennung als auch die Modellierung zur Klassifizierung von Dialog-Tags für Äußerungen, die weiter unter beschrieben werden.

Feature-Extraktion

Die Extraktion und Klassifikation von sprachbasierten Merkmalen (sogenannte Features) des Audiosignals zur automatischen Erkennung von Emotionen umfasst die Extraktion von Low-Level Merkmalen des Sprachsignals sowie die Identifizierung und Extraktion linguistisch motivierter Merkmale basierend auf sprachlichen Lauten, repräsentiert durch Phoneme.

Zur Ermittlung geeigneter Merkmale des Sprachsignals hinsichtlich emotionaler Faktoren gibt es bereits einen umfangreichen Bestand an Arbeiten in der Forschung. Ein exzellenter Überblick ist durch [37] gegeben; darüber hinaus bieten Benchmark-Tests eine Auskunft über die Erkennungsleistung automatischer Erkennungssysteme [139]. Das für dieses Projekt avisierte Ziel ist es ein weitgehend generisches Emotionsmodell zu konstruieren. Zwei Anforderungen werden dabei fokussiert: (1) Die Emotionserkennung soll auf verschiedene Szenarien, nur unter Hinzunahme des Sprecherkontextes, anwendbar sein. (2) Das Resultat der Emotionserkennung ist eine Wahrscheinlichkeitsverteilung über Emotionsklassen, um tiefgreifender Information des Nutzers ableiten zu können.

Die für dieses Vorhaben selektierten Merkmale sowie Merkmalsgruppen sind in Tabelle 2 dargestellt; diese gelten als typische Indikatoren emotionaler Zustände aus der Sprachsignalverarbeitung. Prosodische Merkmale beschreiben die Satzmelodie durch Stimmhöhe (Grundfrequenz – F_0), Intensität (Energy) und Rhythmus. Das Frequenzspektrum einer Äußerung kann in einer kompakten Art und Weise beschrieben werden (MFCCs). Ebenso lässt sich die Konfiguration des Vokaltraktes (bspw. Mundraum, Nasenraum und Zunge) bei stimmhaften Lauten durch Formanten (Verstärkung bestimmter Frequenzbereiche) im Frequenzspektrum charakterisieren. Stimmqualitätsparameter wie Jitter und Shimmer beschreiben wie gehaucht bzw. geräuschbehaftet eine Stimme ist. Zudem wurde in der Forschung bereits der Einfluss von Herzschlag auf den Parameter Jitter untersucht. Wie *rein* eine Stimme ist, zeigt sich mit dem HNR (Harmonics-to-Noise Ratio) Parameter. Nichtlineare TEO (Teager

Energy Operator) basierte Merkmale, insbesondere TEO-CB-Auto-Env (Critical Band based TEO Autocorrelation Envelope Area), wurden in der Forschung zum großen Teil hinsichtlich der automatischen Erkennung von Stress eingesetzt. Dieses auf Stimmen fokussierte Merkmal soll hier ebenso ein Mehrwert für die Erkennung emotionaler Zustände sein. Für die Extraktion fast aller Low-Level Merkmale kam das Audio-Analyse Tool Praat [14] zum Einsatz. Einzige Ausnahme bildet das TEO basierte Merkmal, welches zunächst implementiert werden musste. Hierzu wurde die freie Software-Umgebung von R [153] genutzt.

Messfehler bzw. Artefakte in den Low-Level Konturen ließen sich durch die Anwendung des Simple Moving Average Filters (gleitender Mittelwert) reduzieren. Darüber hinaus erfolgte die Ableitung von Regressionskoeffizienten, um entsprechend das Verhalten der Konturen hinsichtlich Geschwindigkeit (Δ) und Beschleunigung (ΔΔ) zu bestimmen. Um die Merkmale sowie deren abgeleiteten Regressionskoeffizienten in konkrete Merkmalsvektoren zu überführen wurden sogenannte statistische Funktionale abgeleitet (Tabelle 3).

Zusätzlich wurden High-Level Merkmale auf Basis von Phonemen extrahiert. Phoneme bilden auch die Grundlage einer alternativen Repräsentation von sprachlichen Lauten; hierzu wurden Vokale und Konsonanten abgeleitet. Darüber hinaus schafft dies wiederum die Möglichkeit zur Detektion von Silben (vgl. [42]). Um zunächst Phoneme zu ermitteln wurde ein automatischer Phonem-Erkenner auf Basis des Open-Source Frameworks CMU Sphinx 4 [76] implementiert. Eine vorläufige Version wurde bereits in ähnlichen Gebieten eingesetzt [51] Phoneme, Vokale, Konsonanten, Silben und entsprechende Zeitmarken bilden die Grundlage für dynamische und statische Merkmale zur Beschreibung von Rhythmen, welche dem Bereich der Prosodie zugeordnet werden können. Dynamische Merkmale dieser Kategorie berücksichtigen Variationen über die Zeit. Hierzu wurden Funktionale auf Dauer und gefensterte Sprecherraten von Lauten angewendet. Statische Phonem-basierte Merkmale werden hingegen unmittelbar für das gesamte Audio-Segment ermittelt. Dazu zählen: Sprecher- und Artikulationsrate (Phonem- und Silben-basiert), Dauer der Äußerung, gesamte und durchschnittliche Pausendauer, Anzahl von Pausen und das Verhältnis von Pausendauer zu Dauer der Äußerung sowie Anzahl von Pausen zu Anzahl von Phonemen.

Merkmalsgruppe	**Merkmale**	**Funktional**	**#**
Low-Level Merkmale			
Prosodie	F_0+Δ+ΔΔ	A	48
	Intensität+Δ+ΔΔ	A	48
Spektrum	12 MFCCs+Δ+ΔΔ	A	576
	3 Formanten+Δ+ΔΔ	A	144
Stimmqualität	Jitter+Δ+ΔΔ und Shimmer+Δ+ΔΔ	A	96
	HNR+Δ+ΔΔ	A	48
TEO	16 TEO-CB-Auto-Env+Δ+ΔΔ	A	768
Phonem-Level Merkmale			
Prosodie	Dauer (Phoneme, Silben, Vokale, Konsonanten)	B	40
	Sprecherraten (Phoneme, Silben)	C	18
	Statische Phonem-basierte Merkmale	-	10

Tabelle 2: Merkmalsgruppen und extrahierte Merkmale des Sprachsignals für die automatische Erkennung emotionaler Zustände. Gruppen von Funktionalen sind in Tabelle 3 angegeben.

Funktionalgruppe	Funktionale
A	Mittelwert, Standardabweichung, Schiefe, Kurtosis, Minimum, Maximum, Spannweite, Maximum-Mittelwert Differenz, Mittelwert-Minimum Differenz, Position von Minimum, Position von Maximum, Quartil 1, Quartil 2, Quartil 3, Regression β_0 und β_1
B	Mittelwert, Standardabweichung, Minimum, Maximum, Maximum-Mittelwert Differenz, Mittelwert-Minimum Differenz, Position von Minimum, Position von Maximum, Regression β_0 und β_1
C	Standardabweichung, Minimum, Maximum, Maximum-Mittelwert Differenz, Mittelwert-Minimum Differenz, Position von Minimum, Position von Maximum, Regression β_0 und β_1

Tabelle 3: Verwendete Funktionale.

Zur Beurteilung der Erkennungsleistung eines Modells basierend auf den zugrundeliegenden Sprachparametern, wurden Experimente auf verschiedenen Korpora von emotionaler Sprache durchgeführt. Drei populäre Korpora wurden hierzu verwendet:

- *EMO-DB* [19] enthält 535 Sprachaufnahmen von insgesamt 10 Sprechern mit jeweils sieben *gespielten* Emotionen: Anger, Disgust, Fear, Happiness, Sadness, Neutral und Boredom
- *Savee* [62] umfasst 4 Sprecher und insgesamt 480 Sprachaufnahmen mit sieben gespielten Emotionen: Anger, Disgust, Fear, Joy, Sadness, Surprise und Neutral
- *Enterface* [88] umfasst 44 Probanden und insgesamt 1293 Aufnahmen. Die Aufnahmen enthalten sechs induzierte Emotionen: Anger, Disgust, Fear, Joy, Sadness und Surprise

Eine zentrale Frage in der automatischen sprachbasierten Emotionserkennung bezieht sich auf die Frage, welche Aspekte in der Sprache, repräsentiert durch Merkmalsgruppen, für welchen Kontext am geeignetsten sind. Für einen besseren Vergleich der Merkmale wurde zunächst die Sprechervariabilität ausgeschlossen, indem für einen Datensatz jedes Merkmal innerhalb eines Sprecherkontextes Z-transformiert wurde (μ=0, σ=1). Die Relevanz der vier Gruppen Prosodie, Spektrum, Stimmqualität und TEO-basierte Merkmale wurde mittels Entropie-basierter Messungen untersucht. Hierbei kam der *Information Gain* zum Einsatz, welcher ausdrückt, wieviel Information ein Merkmal zur Beschreibung der Klassen beiträgt. Jede Gruppe enthält eine Anhäufung von Merkmalen aufgrund der Anwendung zahlreicher Funktionale. Eine relevante Teilmenge von Merkmalen ist hierbei von Interesse, was durch Merkmalsranking mittels Information Gain erreicht wurde. Die Top 100 Merkmale, welche den höchsten Information Gain für einen gegebenen Datensatz aufzeigen, wurden selektiert. Eine Gruppe wird dann als besonders relevant erachtet, wenn ihr Anteil unter den automatisch selektierten Merkmalen größer ist als ihr Anteil des originalen Merkmalssatzes. Abbildung 11 fasst die automatisch selektierten Merkmale, aufgeteilt in ihre Gruppen, zusammen. Für EmoDB ist die Relevanz von TEO-basierten Merkmalen mit fast 50% Anteil deutlich zu erkennen. Im Gegensatz dazu dominieren prosodische Merkmale Im Fall von Savee und Enterface. Darüber hinaus zeigt sich bei Enterface die Relevanz von Stimmqualitätsparametern.

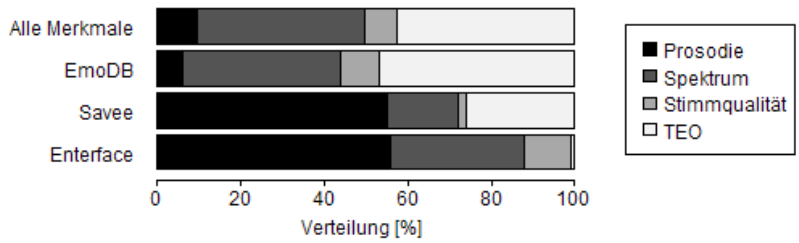


Abbildung 11: Merkmalsrelevanz durch Anteile der Merkmalsgruppen.

Für die automatische Erkennung von Emotionen aus einem gegebenen Sprachsignal, wurde eine entsprechende Prozesskette implementiert (Abbildung 12). In einem ersten Schritt wird das Sprachsignal auf zwei verschieden Arten verarbeitet: (1) eine automatische Spracherkennung detektiert sprachliche Laute (Phoneme) mit entsprechenden Zeitmarken und extrahiert auf dieser Ebene entsprechende Merkmale; (2) Low-Level Merkmale des Zeit- sowie Frequenzbereichs werden aus dem Signal extrahiert. Im nächsten Schritt werden die Merkmalsvektoren aus beiden Verarbeitungsstufen fusioniert und dem Klassifikator (SVM – Support Vector Maschine) übergeben.

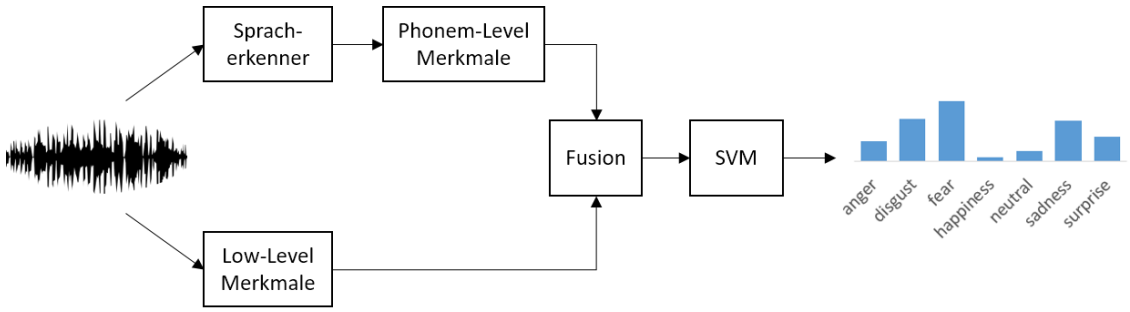


Abbildung 12: Prozesskette zur automatischen Erkennung von Emotionen aus einem Sprachsignal.

In der Regel sind die Erkennungsraten eines Modells nur innerhalb eines Datenbestandes vielversprechend (vgl. [140]). Die Problematik beim Einsatz eines Modells in anderen Kontexten hat vielerlei Gründe: unterschiedliche Aufnahmebedingungen sowie Arten der Indizierung von Emotionen, verschiedene Emotions-Labels und kulturelle Hintergründe. Um jedoch eine generalisierbare Emotionserkennung auf Basis von Audiodaten als Methode für das Gesamtvorhaben zu nutzen, wurde die Diversität emotionaler Daten strategisch durch die Kombination von Datenkorpora untersucht. Dies erfordert zunächst eine annähernd homogene Datengrundlage hinsichtlich der Emotions-Labels für ein überwachtes Lernverfahren. Als Kriterium wurde festgelegt, dass nur Emotionen einbezogen werden, wenn diese von mindestens zwei Korpora zur Verfügung stehen. Tabelle 4 zeigt die für die Untersuchungen genutzte Teilmenge von Emotionen. Es ist zu erkennen, dass „Boredom“ nur im Fall von EmoDB vorliegt; konsequenterweise wurde diese Klasse für Modellierungszwecke ausgeschlossen.

	Anger	Disgust	Fear	Happiness	Sadness	Surprise	Neutral	Boredom
EmoDB	x	x	x	x	x		x	x
Savee	x	x	x	x	x	x	x	
Enterface	x	x	x	x	x	x		

Tabelle 4: Mapping von Emotionsklassen aus verschiedenen Sprachkorpora

Zur Beurteilung der Erkennungsleitung eines Modells basierend auf den zugrundeliegenden Sprachparametern, wurden Experimente unter Verwendung der drei Korpora separat sowie einer Kombination derer durchgeführt. Hierbei ist der Effekt von Phonem-Level, Low-Level und deren Fusion von Interesse. Als Evaluationsstrategie wurde eine *Leave-one-speaker-out* Kreuzvalidierung angewendet, d.h. eine sprecherunabhängige Validierung mit jeweils nur einem Sprecher in den Testdaten, welcher nicht in den Trainingsdaten enthalten ist. In den Experimenten wurden verschiedene Applikations-Szenarien mittels Merkmalsnormalisierung, evaluiert, wobei die zugrundeliegenden Normalisierungsmethoden auf Z-score Standardisierung basieren:

- Trainingsnormalisierung (TN): Normalisierungsparameter werden ausschließlich auf den Trainingsdaten berechnet, wodurch ein Test-sample ad-hoc normalisiert werden kann.

- Sprechernormalisierung (SN): Merkmale werden innerhalb eines Sprecherkontextes normalisiert, um die Sprechervariabilität auszuschließen.
- Korpusnormalisierung (KN): Merkmale werden innerhalb eines Korpuskontextes normalisiert, beispielsweise, um die Variabilität verschiedener Aufnahmebedingungen zu berücksichtigen.

Als Evaluationsmaß kommt der Unweighted Average Recall (UAR) zum Einsatz. Dieses Maß bietet eine wahrheitsgetreue Angabe der Erkennungsleistung, insbesondere wenn ein Ungleichgewicht in der Datenverteilung zwischen den Emotionsklassen vorliegt. Als Lernverfahren wurde eine Support Vector Maschine (SVM) mit polynomialen Kernel eingesetzt. Hierzu wurde der SMO-Algorithmus der Data-mining-Software WEKA 3.8 [49] für das Modelltraining verwendet. Um den optimalen Generalisierungsgrad zu bestimmen, wurden verschiedene Werte des Komplexitätsparameters *C* untersucht: 10^{-5}, 10^{-4}, 10^{-3}, 10^{-2}, 10^{-1} und 1.

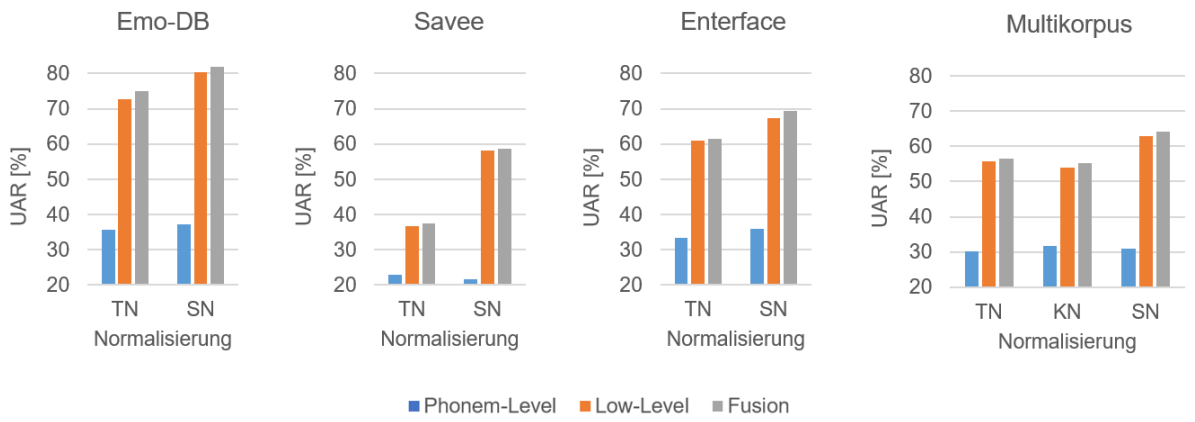


Abbildung 13: UAR (%) Ergebnisse für die automatische Emotionserkennung.

Die UAR Ergebnisse sind in Abbildung 13 zusammengefasst. Die optimalen Werte des Komplexitätsparameters für einzelne Datensätze lagen bei $C=10^{-1}$ (Emo-DB), $C=10^{-2}$ (Savee) und $C=10^{-3}$ (Enterface). Im Fall der Multikorpus-Evaluation lag der optimale Wert bei $C=1$ für Phonem-Level Merkmale und $C=10^{-3}$ für Low-Level sowie fusionierte Merkmale. Wie erwartet ist die Erkennungsleistung auf Basis von rein Phonem-basierten Merkmalen nicht ausreichend, um eine adäquate Emotionserkennung zu modellieren. Es sei allgemein darauf hingewiesen, dass der Einsatz derartiger Merkmale hier als Ergänzung zu Low-Level Merkmalen dient, um die Erkennungsleistung zu optimieren. Dies wird durch geringfügig positive Effekte in den Ergebnissen bestätigt. Hinsichtlich der Normalisierungsmethoden zeigt sich, dass sich die Sprechernormalisierung (SN) positiv auf die Erkennungsleistung im Vergleich zu alternativen Normalisierungsmethoden auswirkt. Jedoch ist dies nicht in allen Realwelt-Szenarien eine Option – eine entsprechende Akquirierung von Daten setzt die Erfassung von einem Sprecherkontext voraus. Bezüglich aller (fusionierter) Merkmale, macht die Sprechernormalisierung nur einen marginalen Unterschied für Emo-DB mit einer absoluten Differenz von 6,9% bzw. Enterface mit einer absoluten Differenz von 7,7%. Im Gegensatz dazu ist der Unterschied bei Savee wesentlich größer mit einer absoluten Differenz von 21,2%. Die besten Ergebnisse für einzelne Korpora sind 81,8% (Emo-DB; State-of-the-Art: 84.6%), 58,6% (Savee; State-of-the-Art: 49,4%) und 69,3% (Enterface; State-of-the-Art: 72.5%). Für näherer Informationen zum aktuellen Forschungsstand siehe [139] [159] [50] [38]. Hinsichtlich der Multikorpus-Evaluation, inklusive 58 verschiedene Sprecher, konnte das beste Ergebnis mit 64.1% ebenfalls durch eine Sprechernormalisierung der fusionierten Merkmale erreicht werden.

Linguistische Merkmale

Neben Merkmalen, welche sich unmittelbar aus dem Audiosignal ableiten lassen, wurden Parameter berücksichtigt, welche sich aus dem linguistischen und semantischen Zusammenhang ergeben. Hierzu wurden Merkmale basierend auf rein textueller Ebene für zwei Arten von Analysen untersucht: (1) die automatische Klassifikation von Dialogakten zur Beschreibung der Kommunikation einer Fluglotsen-Dyade; (2) die automatische Erkennung von Stimmungen in Äußerungen.

Die automatische Erkennung eines Dialogaktes setzt die Analyse einer im Kontext stehenden Äußerung voraus. Hierzu haben sich in der Forschung sogenannte Conditional Random Fields (CRF) [75] etabliert. Der Vorteil dieser probabilistischen Modelle ist, dass sie die komplette Information einer Eingabesequenz berücksichtigen können. Im Zusammenhang der Diskursanalyse ist es somit möglich, vergangene und zukünftige Äußerungen in die Analyse der aktuellen Äußerung einzubeziehen. Maschinelle Lernverfahren setzen in der Regel einheitliche Vektorlängen zur Beschreibung einer Dateninstanz voraus, was eine Transformation der ungleich langen textuellen Äußerungen erfordert. Für diesen Zweck wurde das Konzept von Word2Vec angewendet, welches auf künstliche neuronale Netze basiert. Word2Vec erzeugt aus umfangreichen Textkorpora ein Vektorraum, wobei jedes Wort durch einen Vektor mit konstanter Länge repräsentiert wird. Die Darstellung ganzer Texte (z.B. Satz, Absatz, oder ganze Dokumente) als Vektor kann anschließend durch die Anwendung des arithmetischen Mittels über jeden Index der entsprechenden Wort-Vektoren realisiert werden.

Zur Verifizierung solcher Methoden für die automatische Klassifikation eines Dialogaktes wurden entsprechende Experimente durchgeführt. Die erzeugten Modelle wurden auf Grundlage des SwDA (Switchboard Dialog Act Corpus) [65] evaluiert. Der Datensatz fasst eine Reihe von Telefon-Konversationen in englischer Sprache zusammen. Die Trainingsdaten umfassen 1.115 Konversationen (ca. 197.000 Äußerungen) und 42 verschiedene Dialog-Klassen, welche unterschiedlichen Informationsebenen zugeordnet sind. Die Testdaten hingegen bieten 19 Konversationen mit ca. 5.000 Äußerungen. Das grundlegende Ziel ist die Erzeugung eines Modells zur automatischen Erkennung eines Dialogaktes pro Äußerung. Entsprechend der Trainingsdaten ist dies als ein 42-Klassen-Problem zu betrachten. Merkmalsvektoren für Äußerungen wurden mit 300-dimensionalen Wortvektoren eines vortrainierten Word2Vec Modells (rekurrentes neuronales Netz), welches auf Google News [92] basiert, konstruiert. Für CRF wurde der Kontext-Parameter untersucht, mit K=2, 4 und 6 angrenzenden Äußerungen. Darüber hinaus wurde der CRF-Komplexitätsparameter für Generalisierungszwecke optimiert: $C=10^n$, wobei n = -2, -1, 0, 1 und 2 ist.

Die Ergebnisse sind in (Abbildung 14) zusammengefasst. Es ist deutlich zu erkennen, dass mit zunehmenden Parameter K die Erkennungsleistung steigt. Dies bestätigt die Relevanz kontextbasierter Ansätze. Grundsätzlich zeigt sich auch eine Leistungssteigerung mit zunehmender Komplexität C, d.h., eine geringere Generalisierung des Modells führt zu besseren Ergebnissen. Die beste Leistung wurde mit C=10 und K=6 erzielt (60,19%), mit einer absoluten Differenz von 0,04% zu C=100. Vergleichsweise bessere Erkennungsraten sind in der aktuellen Forschung aufzufinden (71,0% - 79,2%, siehe [73]). Die mit über 10% geringeren Erkennungsraten gegenüber State-of-the-Art Systemen sind darauf zurückzuführen, dass in der hier durchgeführten Evaluation vorrangig eine Abschätzung optimaler Paramater getroffen wurde. Ein offener Punkt hinsichtlich der Optimierung besteht folglich darin, inwieweit eine Erhöhung des Kontextparameters zu einem weiteren Leistungsgewinn beiträgt.

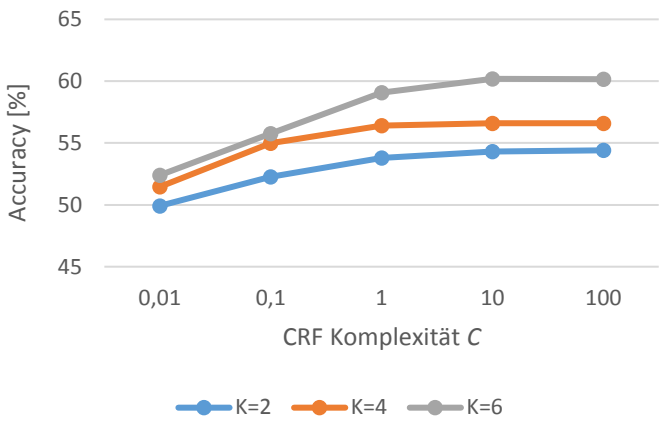


Abbildung 14: Ergebnisse auf den Testdaten des SwDA Korpus. K=Kontext-Parameter.

Der zweite Teil linguistisch motivierter Ansätze bezieht sich auf die sogenannte Sentiment-Analyse, d.h. die automatische Auswertung von Texten, um eine geäußerte Haltung als positive oder negativ einzustufen. Dieses Prinzip wurde im Rahmen dieses Vorhabens zusätzlich auf einer Menge von Emotionen erweitert. Grundsätzlich setzen derartige Methoden umfangreiche Datenkorpora für das Training entsprechender Modelle voraus. Für experimentelle Untersuchungen wurden als Datenquelle die Texte von Twitter genutzt. Beiträge auf Twitter enthalten oft eine hohe Emotionalität. Der Vorteil der mit 140 Zeichen begrenzten Twitter-Beiträgen ist, dass im Normalfall nur eine Emotion beschrieben wird, während sich innerhalb von längeren Texten die Emotionen ändern könnten. Die Suche erfolgte durch Hashtags über die Twitter-API mit Hilfe der twitter4j-Bibliothek [158]. Tabelle 5 fasst die Emotionsklassen, Hashtags und Menge resultierender Dokumente zusammen.

Klasse	Mapping	Hashtags	Anzahl der Dokumente
Joy	Positiv	#joy, #happy	4.179
Anger	Negativ	#anger, #angry, #hate	3.688
Fear	Negativ	#fear, #scared	2.750
Sadness	Negativ	#sadness, #sad	1.976
Disgust	Negativ	#disgust, #disgusting, #gross	3.308
Surprise	-	#surprise, #surprised	2.947
Neutral	Neutral	#news	3.705

Tabelle 5: Erzeugter Datensatz auf Basis von Twitter zur Sentiment-Analyse

Für die Umsetzung der Sentiment-Analyse wurde ein Java-basiertes Framework implementiert, welches optimale Kombinationen von Merkmalen und Methoden evaluiert. Die Komponenten des Frameworks sind in (Abbildung 15) illustriert. In der ersten Stufe findet die Vorverarbeitung von textuellen Daten statt. Dabei werden die Inhalte der Dokumente in eine abstrahierte, bearbeitbare Form konvertiert. Anschließend erfolgt für je ein Dokument die Extraktion von Merkmalen, welche zu einem Merkmalsvektor zusammengefasst werden. Folgende Merkmale werden im Framework für Dokumente unterstützt:

- Bag-of-N-Grams (BoW, Bo2G, Bo3G): Vorkommen von Vokabular und Wortkombinationen
- WordNet [93]: Hypernyme (Sammelbegriffe) und Synonyme für alle Wörter
- LDA (Latent Dirichlet Allocation) [13]: Zuordnung von Dokumenten zu abstrakten Kategorien
- Part of Speech: Beschreibung hinsichtlich der Wortarten (Substantiv, Verb oder Adjektiv)
- Schlagwörter [15] Vordefinierte Valenz-Werte in Wortlisten dienen als Merkmal

- Satzzeichen

Zwei Verarbeitungsstrategien wurden hierbei implementiert: (1) Die extrahierten Merkmale werden direkt zu einem *großen* Merkmalsvektor zusammengefasst (Early-Fusion) und anschließend einer Klassifikationsmethode übergeben. (2) Für je ein Merkmal existiert ein eigener Klassifikator. Das Ergebnis jedes Klassifikators wird anschließend verknüpft (Late-Fusion). Die folgenden Late-Fusion Strategien wurden implementiert: Maximum, Durchschnitt und Produkt von Konfidenzwerten sowie Mehrheitsentscheid der Klassifikatoren.

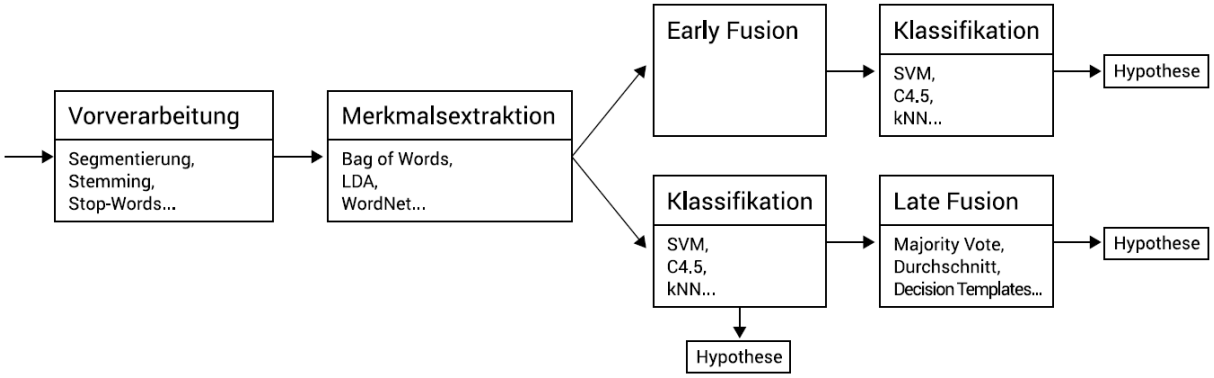


Abbildung 15: Framework für Text-basierte Emotionserkennung.

In Experimenten wurden verschiedene Kombinationen von Merkmalen sowie Fusionsmethoden untersucht. Neben der Untersuchung einzelner Merkmale war die Leistung aller Merkmalskombinationen mittels Early-Fusion sowie Late-Fusion von Interesse. Als Klassifikationsmethode kamen ausschließlich Support-Vektor Maschinen zum Einsatz. Hierzu wurde die Implementierung der LibLinear-Bibliothek [40] verwendet. Alle Experimente wurden mit einer 10-fachen stratifizierten Kreuzvalidierung ausgeführt, so dass jede der 10 Teilmengen eine annähernd gleiche Verteilung besitzt. Es wurde der Unweighted Average Recall (UAR) als Evaluationsmaß verwendet, da eine ungleiche Verteilung zwischen den Klassen vorliegt.

Tabelle 6 zeigt die Ergebnisse der Kategorien bester Einzel-Klassifikator, beste Early-Fusion und beste Late-Fusion für das 3- und 7-Klassenproblem. Bezüglich des 3-Klassenproblems konnte eine Verbesserung durch beide Fusionsmethoden gegenüber dem besten Einzel-Klassifikator festgestellt werden. Das beste Ergebnis mit 80,34% wurde durch Early Fusion der Merkmale Bag-of-Words (BoW) und Bag-of-Trigramms (Bo3G) erreicht. Interessanterweise bringen im Fall des 7-Klassenproblems nicht alle Fusionsmethoden eine Verbesserung: Early-Fusion liegt mit einer absoluten Differenz von 0,09% hinter dem Einzel-Klassifikator (62,85%). Das beste Ergebnis wurde durch Late-Fusion mittels Durchschnittsberechnung ermittelt (63,87%). Dies entspricht einer Verbesserung von 1,02% gegenüber dem Einzel-Klassifikator und 1,11% gegenüber Early-Fusion. Insgesamt zeigen die Ergebnisse für beide Klassifikationsprobleme, dass Bag-of-Ngramms, insbesondere Bag-of-Words (BoW), in allen Szenarien als favorisiertes Merkmal betrachtet werden kann.

Methode	Merkmale	UAR
3 Klassen		
Einzel-Klassifikator	BoW	79,66%
Early-Fusion	BoW + Bo3G	80,34%
Late Fusion (Maximum)	BoW + Bo2G	79,88%
7 Klassen		
Einzel-Klassifikator	BoW	62,85%
Early-Fusion	BoW + Bo3G	62,76%
Late Fusion (Durchschnitt)	BoW + Bo2G + Bo3G + WordNet	63,87%

Tabelle 6: Beste UAR (%) Ergebnisse verschiedener Methoden für 3 und 7 Klassen.

Spracherkennung

Die Extraktion linguistischer Merkmale im laufenden Fluglotsen-Betrieb setzt die automatische Transkription von Äußerungen voraus. Da nicht nur Funksprüche, die auf Englisch getätigt werden, betrachtet werden, ist auch die Muttersprache des Fluglotsen – in diesem Fall Deutsch – zu berücksichtigen. Drei verschiedene Szenarien sind grundsätzlich zu unterscheiden: (1) Zwei Fluglotsen stehen im Dialog, wobei jeder Einzelne durch ein separates Mikrofon erfasst wird; (2) Der Fluglotse sendet einen Funkspruch; (3) Gesamtheit des Funkspruches und der Kommunikation zu einem anderen Fluglotsen (die Kombination aus (1) und (2)).

Aus technischer Sicht gibt es verschiedene Strategien für die Implementierung multilingualer Spracherkenner. In der Forschung zeigen sich vielversprechende Ansätze in der Nutzung sprachenübergreifender Modelle [141] d.h., Äußerungen verschiedener Sprachen werden durch einen Spracherkenner ad-hoc erkannt. Dieser Ansatz wurde auch in diesem Vorhaben verfolgt. Dies erfordert eine zielgerichtete Entwicklung der folgenden drei Wissensquellen eines Spracherkenners: Aussprachewörterbuch, Akustikmodell und Sprachmodell. Zudem benötigen die ersten zwei genannten Quellen ein einheitlich phonetisches System.

Bevor das Fluglotsenszenario genauer beleuchtet wird, ist zunächst der generalisierte Fall einer bilingualen Spracherkennung umzusetzen. Zur Schaffung eines einheitlichen phonetischen Systems wurden für die deutsche und englische Sprache zunächst jeweils Aussprachewörterbücher erstellt. Als Datengrundlage für das deutsche Wörterbuch wurde der Sprachdatenkorpus *German open source corpus for distant speech recognition* [120] sowie der Textdatenkorpus von *Europarl* [72] verwendet. Nach Zusammenführen der textuellen Daten ergibt sich ein Vokabular von ca. 311.000 Wörtern. Im Fall des englischen Aussprachewörterbuches bietet sich das Vokabular vom Sprachdaten-Korpus LibriSpeech [106] an; hierdurch wurde ein Vokabular von ca. 260.000 Wörtern extrahiert.

Der nächste Schritt beinhaltet die Erzeugung der phonetischen Transkription des Vokabulars. Für diesen Zweck wurden die BAS Web-Services [126] zur Erzeugung der Aussprache mittels des phonetischen Alphabets SAMPA genutzt. In der Summe ergeben sich 66 Phoneme für Deutsch und 41 Phoneme für Englisch. Um die Komplexität bei der Beschreibung durch Phoneme zu reduzieren, bspw. durch Ausschluss enger Verbindungen (Affrikate) wie *pf* und *tS*, wurde der Satz an Phonemen für Deutsch auf 42 und für Englisch auf 38 reduziert. Durch Zusammenführen beider Phonemsysteme ergibt sich, wie in Tabelle 7, dargestellt, ein gesamtheitliches System für beide Sprachen.

Phonemsystem Deutsch-Englisch
2:, 6, 9[D], @, a, a:[D], aI, aU, b, C[D], d, D, E, e:[D], E:[D], eI, f, g, h, I, i:, j, k, l, m, n, N, O, o:, OY, p, r[D], R[E], s, S, t, T[E], U, u:, v, w[E], x[D], Y[D], y:[D], z, Z, SIL

Tabelle 7: Bilinguales Phonemsystem für Deutsch und Englisch in der SAMPA Notation. [D]: ausschließlich Deutsch, [E]: ausschließlich Englisch. Für nähere Informationen zu den gelisteten Phonemen, siehe [160].

Für die Konstruktion eines Sprachmodells im bilingualen Kontextes wurde zunächst für die jeweilige Sprache ein eigenes Sprachmodell berücksichtigt. Darauf erfolgt die Interpolation beider Modelle zu einem bilingualen Sprachmodell. Die hierzu eingesetzten Modelle sind sogenannte Trigramm-Modelle, welche dadurch charakterisiert sind, dass sie Wahrscheinlichkeiten von drei aufeinanderfolgenden Wörtern schätzen. Für Englisch fiel die Wahl auf ein vortrainiertes Sprachmodell von LibriSpeech mit der Bezeichnung „3-gram.pruned.3e-7“ [106] Für Deutsch wurden die Textdaten von Europarl [72] verwendet und mit Hilfe des SRILM Toolkits [150] ein Sprachmodell erzeugt.

Für das Training akustischer Modelle für Englisch kam ebenfalls der Korpus LibriSpeech zum Einsatz. Hierfür wurde ein Teilkorpus verwendet („train-clean-100“) welcher ca. 100 Stunden Audioaufnahmen umfasst. Für Deutsch wurden die Sprachdaten des *„German open source corpus for distant speech recognition“* verwendet. Dieser umfasst ca. 33 Stunden Audioaufnahmen. Beide Sprachdatenkorpora enthalten zudem einen Entwicklungs- und Testdatensatz. Akustikmodelltraining sowie Spracherkennung wurde mittels der Open-Source Software CMU Sphinx [76] durchgeführt.

In vorläufigen experimentellen Untersuchungen wurden zunächst zwei Konfigurationen (16 und 32 Gaußsche Komponenten) hinsichtlich der Hidden-Markov-Modell (HMM) basierten Akustikmodellierung untersucht. Für beide Sprachen ergaben sich jeweils die besten Ergebnisse mittels 32 Komponenten: für Deutsch 44.6% (46,74% auf Entwicklungsdaten) und für Englisch 25,42% (25,41% auf Entwicklungsdaten). Im Weiteren wurden die Sprachen bzgl. der Entwicklungsdaten und Testdaten kombiniert, um den bilingualen Fall zu evaluieren. Die Spracherkennung wurde entsprechend folgendermaßen ausgerichtet: Es wurden die sprachübergreifenden Wörterbücher kombiniert und das bilinguale Akustikmodell auf Basis der zusammengeführten Trainingsdaten mit 32 Gaußschen Komponenten trainiert. Für die Evaluation wurden drei verschiedene Interpolationsgewichte für das bilinguale Sprachmodell untersucht, wobei das Gewicht sich jeweils auf die deutsche Sprache bezieht. Die Ergebnisse sind in Tabelle 8 dargestellt. Das beste Ergebnis wurde sowohl auf den Entwicklungs- also auch auf den Testdaten mit dem Gewicht λ=0.25 erreicht.

λ	Entwicklungsdaten	Testdaten
0,25	36,31%	36,33%
0,50	37,16%	36,83%
0,75	39,28%	38,49%

Tabelle 8: Wortfehlerrate für bilinguale Spracherkennung. λ: Interpolationsgewicht für Deutsch.

Der Einsatz des bilingualen Spracherkenners für den Bereich der Fliegersprache erfordert eine Anpassung des Sprachmodells sowie des Aussprachewörterbuches. Relevante Texte für das Sprachmodell der Fliegersprache konnte aus dem "Air Traffic Control Simulation Speech Corpus" ATCOSIM [53] gewonnen werden. Das Training des Sprachmodells wurde mit dem bereits verwendeten SRILM Toolkit durchgeführt. Anschließend erfolgte eine Adaption des deutschen Sprachmodells unter Verwendung der drei Interpolationsgewichte Gewichte 0,25, 0,5 und 0,75. Darüber hinaus bildet der ATCOSIM Korpus die Grundlage für das Vokabular der Fliegersprache. Mit Hilfe der BAS Web-Services [126] wurde

das Wörterbuch der Fliegersprache phonetisch transkribiert; der Satz an Phonemen wurde wie oben beschrieben reduziert. Der nächste Schritt bestand darin, die Phoneme im neu entstandenen Aussprachewörterbuch anzupassen. Hauptsächlich betrifft diese Anpassung die geäußerten Buchstaben und die Ziffern für die Flug-Koordinaten [58].

Für die Evaluation im Bereich der Fliegersprache liegen die transkribierten Daten aus AP 5.2 zugrunde. Neben der Verwendung trainierter Modelle, wurden zusätzlich Modelladaptionen zur Optimierung durchgeführt. Hierzu wurde der Fluglotsendatensatz aus AP 5.2 partitioniert, wobei ein Teil zum Adaptieren und ein Teil zum Testen genutzt wurde. Die erste Adaption bezieht sich auf die sprecherunabhängige Akustikmodell- und die Sprachmodelladaption. Die Teilung erfolgte so, dass am Ende zwei möglichst ausgeglichene Partitionen hinsichtlich des Geschlechts, Art der Kommunikation und Dauer vorliegen. Eine zusätzliche Adaption soll Aufschluss über den Effekt der Sprecher-Abhängigkeit geben. Hierzu wurde die Menge an Sprachdaten je Sprecher mit 50% für die Adaption und 50% für den Test partitioniert. Für die Adaption des Akustikmodells wurde die maximum a posteriori (MAP) Methode genutzt. Im Fall der Adaption des zugrundeliegenden Sprachmodells wurde zunächst auf Basis der textuellen Daten des Fluglotsendatensatz aus AP 5.2 ein Trigramm-Model trainiert und anschließend mit dem gegebenen Sprachmodell gewichtet interpoliert.

Tabelle 9 fasst die experimentellen Ergebnisse zusammen. Die Tabelle zeigt, dass sich eine Adaption positiv auf die Erkennungsraten auswirkt. Zudem ist ersichtlich, dass die sprecherabhängige Adaption im Vergleich zur sprecherunabhängigen Adaption eine deutlich geringere Wortfehlerrate für Funksprüche liefert. Für Nichtfunksprüche kann eine geringfügige Verbesserung festgestellt werden. Die Verbesserungen sind auf die sprecherabhängige Adaption des akustischen Modells zurückzuführen. Andererseits dürfte auch der veränderte Kontext durch die Adaption des Sprachmodells zu den Verbesserungen beigetragen haben. Nichtsdestotrotz ist die Wortfehlerrate für Nichtfunksprüche sehr hoch. Dies zeigt, dass die Spracherkennung für Nichtfunksprüche zu fehlerbehaftet für den Einsatz in der Praxis ist. Grundsätzlich kann davon ausgegangen werden, dass die zu hohe Fehlerrate mit der Problematik von Spontansprache im Zusammenhang steht – das frei formulierte, unvorbereitete Sprechen im Alltag, welches nach wie vor ein kritischer Punkt für automatische Spracherkennungssysteme ist.

λ	**Radio**	**No Radio**	**Radio + No Radio**
0,25	70,86%	96,73%	84,50%
0,50	72,71%	96,05%	85,12%
0,75	75,49%	95,61%	86,19%
AM + SM Adaption (su)			
λ_1=0,5, λ_2=0,25	43,30%	89,10%	67,93%
λ_1=0,5, λ_2=0,50	43,87%	89,33%	68,35%
λ_1=0,5, λ_2=0,75	45,06%	90,30%	69,42%
AM + SM Adaption (sa)			
λ_1=0,5, λ_2=0,25	29,64%	85,12%	58,65%
λ_1=0,5, λ_2=0,50	29,76%	85,67%	59,00%
λ_1=0,5, λ_2=0,75	30,69%	86,42%	59,84%

Tabelle 9: Wortfehlerrate der bilingualen Spracherkennung für die Sprache aus der Fluglotsen-Simulation. λ/λ_1: Interpolationsgewicht für Deutsch bei Adaption mit der Fliegersprache (ATCOSIM). λ_2: Interpolationsgewicht für Deutsch + Fliegersprache (ATCOSIM) bei Adaption mit der Sprache aus der Fluglotsen-Simulation. Radio: Funksprüche; AM: Akustikmodell; SM: Sprachmodell; su: Sprecherunabhängig; sa: Sprecherabhängig.

Abgleich und Integration

Grundsätzlich existieren verschiedene Möglichkeiten zur Integration der Ergebnisse der Audioanalyse in das Gesamtsystem. Die generierten Informationen der Audioanalyse sind vielschichtig - sie reichen von Low-Level Merkmalen über interpretierte Diskursanalysen bis hin zu klassifizierten Emotionen. Der Abgleich mit einem Gesamtsystem ist jedoch notwendig, da sich Anforderungen bezüglich der Funktionsweise der multimodalen Fusion ergeben können. In der Regel ist zwischen den Arten der Modalitätsinformation abzuwägen [22]: Die Überführung finaler Klassifikationsergebnisse reduziert unter Umständen den Informationsgehalt einer Modalität. Eine große Menge an Low-Level Parametern kann dagegen die Komplexität bei der Informationsverarbeitung erhöhen.

Um zu gewährleisten, dass Ergebnisse der Audio-Analyse für jede strategische Vorgehensweise eines Gesamtsystems nutzbar sind, wurden mehrere Integrationsmöglichkeiten konzipiert:

- Klassifikationsergebnisse: eine vorhergesagte Klasse wird unmittelbar übertragen (bspw. als nomineller Wert).
- Verteilung in den Klassen: Das Ergebnis ist eine Wahrscheinlichkeitsverteilung über die Klassen, ermittelt durch den Klassifikator.
- Merkmale: ein Satz an extrahierten Merkmalen oder eine durch Merkmalsselektion (bspw. durch filterbasierte Verfahren) reduzierte Menge an Merkmalen wird als Merkmalsvektor bereitgestellt.
- Kombinatorischer Ansatz: bei mehreren Informationskanälen der Audio-Analyse können Varianten kombiniert werden, bspw. Verteilung von Emotionen und Merkmale der Diskursanalyse.

Aus technischer Sicht bietet das experimentelle Framework zur automatischen Erkennung von Emotionen aus Text bereits Komponenten für die Verknüpfung von Merkmalen sowie eine Auswertung auf Basis verschiedener Klassifikationsergebnisse. Diese Komponenten können ebenso für die automatische Klassifikation basierend auf Audiosignalen angewendet werden. Damit ergibt sich die Möglichkeit zur Integration aller Ergebnisse der Audioanalyse auf verschiedene Verarbeitungsebenen.

1.4 Psycho-physiologische Messdaten (AP 6.1 – 6.4)

Auf Basis eines ausführlichen Literaturreviews wurden die Erhebungsinstrumente zur Detektion des aktuellen emotionalen bzw. kognitiven Zustands anhand psycho-physiologischer Daten sowie der relevanten Persönlichkeitsmerkmale für die beteiligten Kanäle ausgewählt. Anschließend erfolgte eine Anpassung der technischen Geräte an die beim Projektpartner vorherrschenden lokalen Begebenheiten der Simulationsumgebung, in der die Datenerhebungen stattfinden sollten. Das Mess-Setup wurde mittels eines Pre-Tests evaluiert und mit den beteiligten Partnerprofessuren abgestimmt. Darüber hinaus wurde eine Strategie für die Datenaufbereitung sowie die anschließende Auswertung entwickelt. Dazu wurde ein weiterer Pre-Test durchgeführt, welcher auf Basis eines kognitiv anspruchsvollen Videospiels mit multiplen zu koordinierenden Elementen ein Analogon zur typischen Centerlotsen-Arbeit darstellt und damit eine nähere Betrachtung der potenziell relevanten psycho-physiologischen Parameter ermöglichte, wobei sich vor allem die Hautleitfähigkeit sowie die Pupillenerweiterung als wichtigste Indikatoren zeigten Abbildung 16(Abbildung 16).

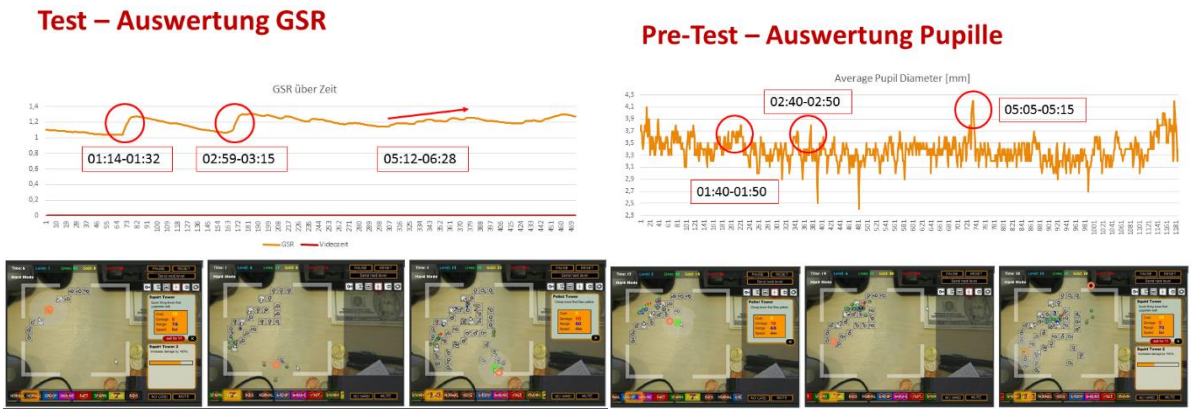


Abbildung 16: Pre-Test Hautleitfähigkeit und Pupillenerweiterung.

Die Sensorik wurde im Anschluss an die Pre-Test-Validierung in Kooperation mit der DFS sowie den übrigen datenerhebenden Professuren in Langen an einem gesonderten Experimentalbereich aufgebaut. In diesem Zusammenhang wurde im Verlauf der darauffolgenden Datenerhebung auch ein Wechsel der Biofeedback-Geräte vorgenommen, da die Verkabelung von Fingern zur Abnahme der Hautleitfähigkeit sowie des Blutvolumenpulses der Nexus-10-Sensoren (Mind Media) eine zu starke Einschränkung der Arbeitsfähigkeit von Fluglotsen zur Folge hatte. Deswegen wurde im weiteren Verlauf des Projekts auf ein Wearable Design der Firma Empatica vertraut. Beide Geräte ermöglichen eine kontinuierliche Aufzeichnung der relevanten psycho-physiologischen Daten für eine anschließende Annotation (Abbildung 17).

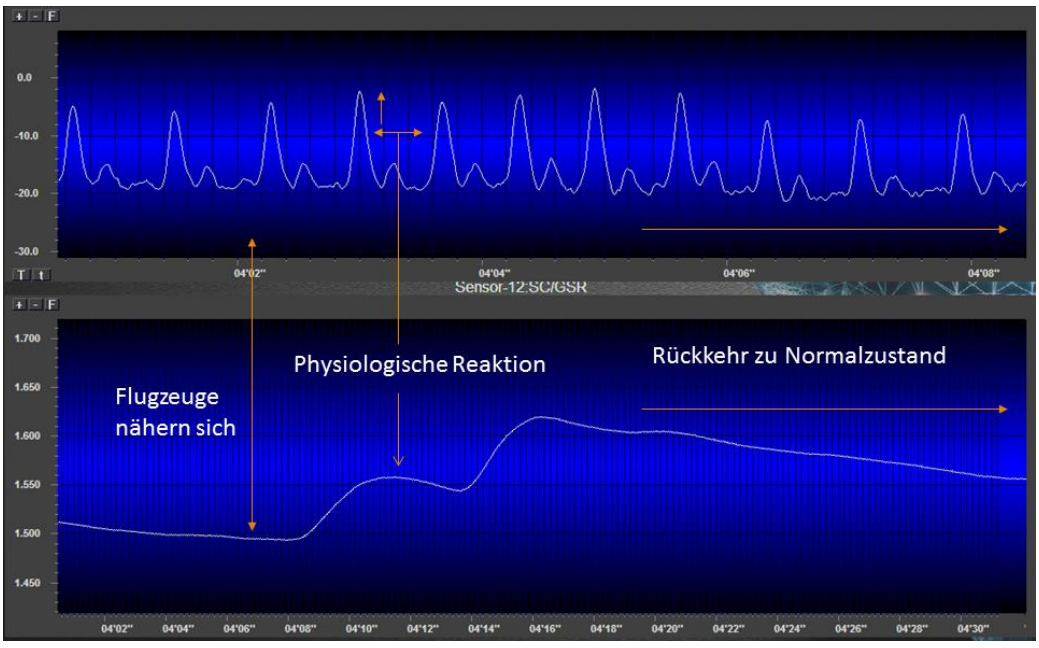


Abbildung 17: Prototypische Zuordnung eines Events zu psycho-physiologischen Erregungspotentialen

Die Pupillendilatation und das Blickverhalten (Abbildung 18) wurde mittels einer mobilen Eye-Tracking-Brille der Firma SMI aufgezeichnet.

Abbildung 18: Blickbewegungserfassung des Radarlotsen.

Darüber hinaus wurden mittels geeigneter psychologischer Tests (KLT-R, FAIR-2, BigFive, MDBF) die Aufmerksamkeit, die Konzentrationsfähigkeit sowie die emotionale Befindlichkeit der Lotsen gemessen. Dabei zeigte sich, dass Fluglotsen über eine außergewöhnlich hohe Fähigkeit zur aufmerksamen und konzentrierten Arbeit verfügen, welche in seiner Ausprägung die auf eine Normalbevölkerung geeichten Tests an ihre Grenzen führt (Abbildung 19).

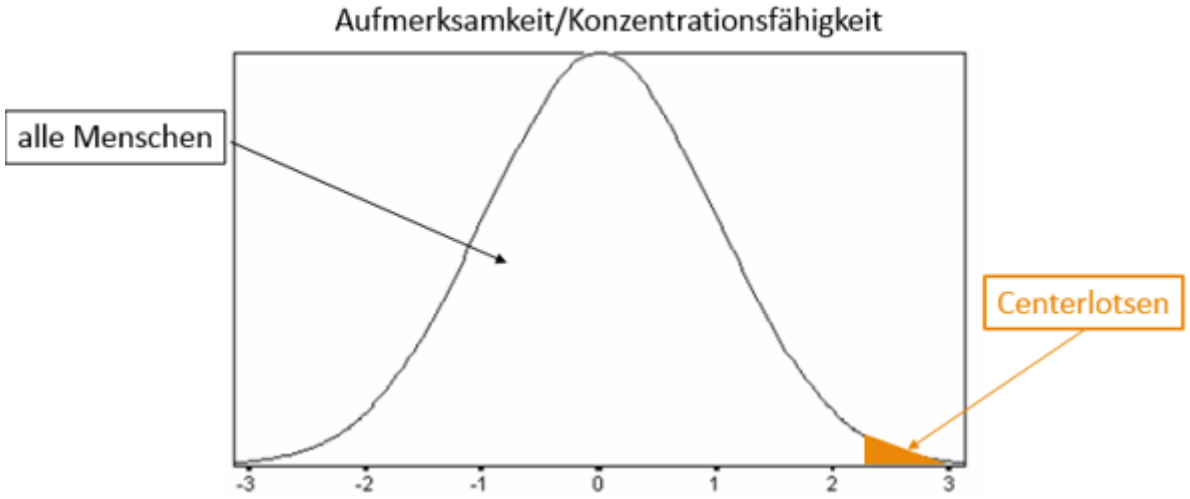


Abbildung 19: Aufmerksamkeit und Konzentrationsfähigkeit von Fluglotsen.

Die Zusammenführung von beobachtbarem Verhalten der Fluglotsen und den Daten der biophysiologischen Erhebung ist ein essentieller Bestandteil der Modellerstellung. Dabei wurden die Blickbewegungen der Fluglotsen durch ein Auswertungsverfahren namens „Semantic Gaze Mapping“ analysiert. Dadurch ist es nun möglich, die Verteilung der Aufmerksamkeit auf unterschiedliche Arbeitsbereiche während einer Lotsenaktivität präzise zu bestimmen (Abbildung 20). Seitens der DFS wurden Logfiles der Simulationen zur Verfügung gestellt, die objektive Daten (u. a. Anzahl der Flugzeuge im Sektor, Anzahl der Funksprüche zwischen Lotsen und Piloten) liefern. Diese Daten flossen in die Modellerstellung ein (siehe 1.6).

Area of Interest	Net Dwell Time Total [ms]
unspezifisch	179268
Blick zu Nachbar	45389
rechter Bildschirm oben	15628,3
linker Bildschirm oben	38670,5
Ista	55241,5
Flugstreifen	1084921
Radarschirm	1682911,6
White Space	8776,6

Abbildung 20: Beispiel einer Area-of-Interest-Analyse der Fluglotsenaufmerksamkeitszuweisung.

1.5 Datengenerierung und Sensordatenfusion (AP 2, AP 16.1, 16.2)

Im Verlaufe des ersten Projektjahres wurden insgesamt drei Simulationstage mit insgesamt sechs Simulationsläufen (jeweils eine Stunde) im Zentrum für Forschung und Entwicklung bei der DFS in Langen realisiert. Ziel der Simulationsläufe war es, detaillierte Daten zu konkreten Arbeitsabläufen eines prototypischen Lotsenteams (Executive und Planner) zu erhalten sowie gleichzeitig Messungen zum kognitiven und emotionalen Erleben und Befinden vorzunehmen. Konkret wurden während der Simulationsläufe folgende Messungen durchgeführt:

1. Videoaufzeichnung der Gesichter der Fluglotsen per HD-Webcam für die weitere Auswertung zur Erkennung mimischer Emotionsausdrücke per FACS (siehe 1.2).
2. Videoaufzeichnung der Fluglotsen per stereoskopischer Kamera für die weitere Auswertung von Posen und Gesten (siehe 1.1).
3. Videoaufzeichnung des kompletten Lotsenarbeitsplatzes per HD-Webcam zur Identifizierung von Interaktionen zwischen den Lotsen (siehe 1.1).
4. Messung des Blickbewegungsverhaltens inklusiver Erfassung der Pupillendilatation per Blickbewegungsbrille zur weiteren Auswertung der kognitiven Belastung sowie der Interaktion mit dem Arbeitsplatz (siehe 1.4).
5. Messung des zeitlichen Verlaufs der Hautleitfähigkeit und der Herzrate zur weiteren Diagnostik des kognitiven und emotionalen Erlebens (siehe 1.4).
6. Audioaufzeichnung der verbalen Kommunikation der Center-Lotsen zur weiteren Diagnostik des kognitiven und emotionalen Erlebens (siehe 1.3).
7. Erfassung der subjektiven Belastung (ISA) und des subjektiven Situationsbewusstseins im zeitlichen Verlauf (Urteil der Lotsen im Abstand von fünf Minuten) (siehe 1.4 und 1.6).

Die aufgenommenen Daten wurden hinsichtlich Korrelationen untersucht und statistisch analysiert. Dabei wurden verschiedene Events (z. B. Kognitive Belastung hoch, Stimme erregt, Pilot spricht zu Fluglotse) in den jeweiligen Kanälen (Video, Audio, Psychophysiologie) markiert bzw. definiert. Diese annotierten Ereignisse wurden zur Weiterverarbeitung an die Sensordatenfusion weitergegeben, welche anschließend mittels Emotionsmodell aus allen Teilzuständen eine Gesamtemotion generiert.

Für die Zusammenführung der Daten wurden mehrere Sensordatenfusionsansätze [5][23][24][71] erwogen. Für die Umsetzung ausgewählt wurde der Ansatz "decision level multimodal fusion". Hauptgründe für diese Wahl waren die vielen verschiedenen multimodalen Kanäle (FACS, Sprache, physiologische Daten, Posen) sowie die Tatsache, dass die Expertise zur Auswertung der einzelnen Kanäle bei den jeweiligen Projektpartnern lagen (vgl. Abbildung 21). Um eine effektive Zusammenarbeit zu gewährleisten, wurde jeder Kanal von der jeweiligen verantwortlichen Professur hinsichtlich der auftretenden Emotionen und kognitive Belastung bzw. deren Indikatoren sowie nach der Qualität der Kommunikation der Fluglotsen analysiert. Diese Einzelauswertungen wurden anschließend in der Sensordatenfusion zu einem ganzheitlichen Kognitions- und Emotionszustand des Fluglotsen zusammengefasst. Abbildung 21 stellt das entwickelte Konzept der Sensordatenfusion grafisch dar.

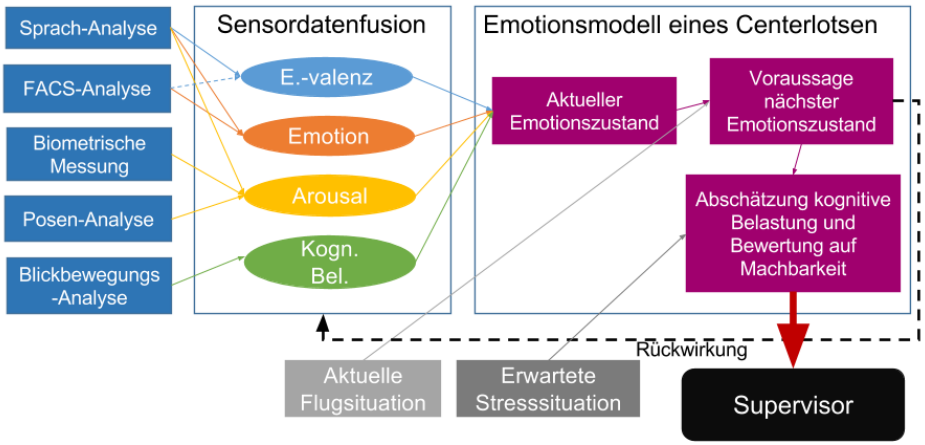


Abbildung 21: Darstellung der Verarbeitung der Ergebnisse in der Sensordatenfusion und deren Weiterverarbeitung im Emotionsmodell der Centerlotsen

Die verschiedenen Kanäle wurden hinsichtlich ihrer Aussagekraft über die Parameter Emotionsvalenz, Emotionsklasse (z. B. Ärger, Wut, Überraschung), emotionale Erregung (Arousal) und kognitive Belastung untersucht. Dabei können nicht alle Kanäle Aussagen über alle Parameter geben. So kann die FACS-Analyse nur Daten zum emotionalen Zustand der Probanden liefern, biometrische Messungen und die Posenanalyse nur Daten zu Arousal und kognitiver Belastung. Die Aussagen der Kanäle wurden in den jeweiligen Parameterkategorien fusioniert mit dem Ziel, anhand dieser Daten einen konsistenten Gesamtzustand eines Fluglotsen zu berechnen (siehe Abbildung 22).

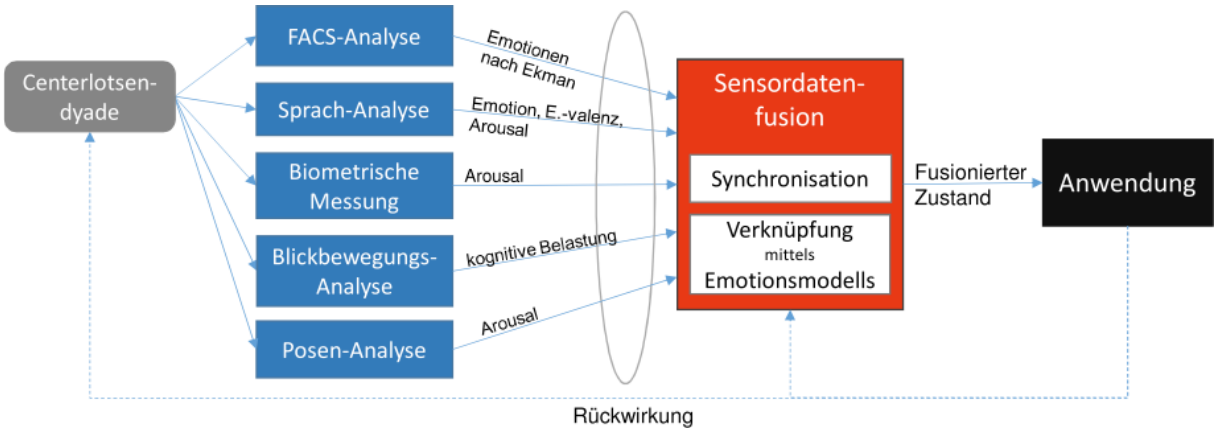


Abbildung 22: Konzeption der verschiedenen Eingabeparameter in die Sensordatenfusion

Die Auswertung der erhobenen Daten offenbarte zwei Probleme der gewählten Vorgehensweise:

- Die erhobenen Daten zeigten nur sehr geringe Varianzen in den untersuchten Variablen, sodass es auf Kanalebene nicht möglich war, eine verlässliche Kategorisierung der Emotionsklasse oder Bewertung des Arousals vorzunehmen. Die geringe Expressivität der Messdaten ist damit zu erklären, dass die Simulationsevents der DFS, die zur Datenaufnahme genutzt wurden, nicht vorsehen, starke Belastungssituationen für die Fluglotsen zu erzeugen, sondern eher einen normalen Arbeitsbetrieb abbilden.

- Außerdem gab es Probleme bei der Extraktion der zugehörigen Situationsvariablen (welche Flugzeuge befanden sich wo, wie viele gab es auf dem Schirm, usw.), und auch die objektive Einschätzung des Schwierigkeitsgrads anhand der gegebenen Daten erschien Experten schwierig.

Aus diesen Gründen wurden für unsere Modellbildung eigene Experimente durchgeführt. Hierzu wurde eine eigens entwickelte Fluglotsen-Simulation in Unity-3D programmiert, welche die Fluglotsenaufgaben so einfach wie möglich darstellt und gleichzeitig alle für die Modellierung notwendigen Parameter automatisch in einem Logfile abspeichern konnte.

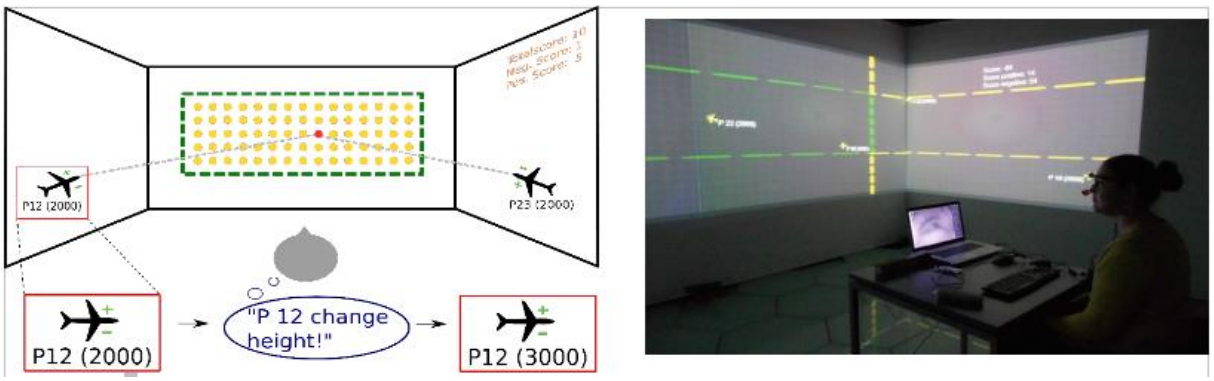


Abbildung 23: Darstellung der entwickelten Simulation (links schematisch, rechts im realen Einsatz).

Die Fluglotsen-Simulation beinhaltete die Darstellung eines dreigeteilten Radar-ähnlichen Bildschirms. Das mittlere Radarbild zeigte einen grün-umrandeten Flugsektor, für welchen der jeweilige Experi-

mentteilnehmer – ähnlich wie in der realen Fluglotsenarbeit – verantwortlich war (siehe auch Abbildung 23). Innerhalb der Simulation erschienen mehrere Flugzeugpaare von oben, unten, links und rechts in einem vorgegebenen und durch Parameter variabel zusetzenden Zeitabstand.

Das automatische Log-System speicherte während der verschiedenen Experimentsessions sowohl die Position, Zielpunkte und Höhen der Flugzeuge als auch bestimmte Ereignisse (Erscheinen, Kollisionen und Verschwinden von Flugzeugen) und deren Zeitpunkte. Außerdem wurde mit Hilfe eines Punktesystems ein individueller Score ermittelt, welcher genutzt werden konnte, um die Leistung der Probanden untereinander zu vergleichen.

Drei Experimente, die im ursprünglichen Arbeitsplan nicht vorgesehen waren, wurden innerhalb des Projektes durchgeführt:

Experiment 1 in Barcelona: September-November 2015

Es wurden Daten von 25 Probanden (Mittelwert Alter = 28,1; SD = 5,7; 64,0% männlich) mit mehrheitlich keinen Erfahrungen von Fluglotsenaufgaben aufgezeichnet. Das Experiment untersuchte, inwiefern eine Pupillenerweiterung und eine Erhöhung des Arousals valide Indikatoren für die Einschätzung der Arbeitsbeanspruchung sind.

Das 2 (Schwierigkeitsgrad der Aufgabe) x 3 (Ereignisse) Innersubjekt-Experimentdesign beinhaltete die Erfassung von Persönlichkeitsmerkmalen (durch BigFive-Fragebogen erfasst), die momentane Befindlichkeit der Probanden (durch den englischen Mehrdimensionalen Befindlichkeitsfragebogen MDMQ erfasst) und die verschiedenen Pupillenerweiterungs- und EDA-Daten aller Probanden.

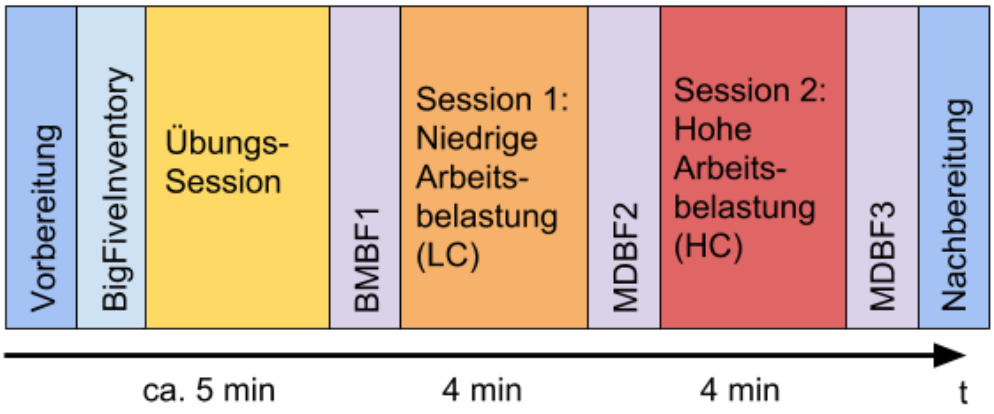


Abbildung 24: Abbildung der Experimentprozedur.

Leider konnten die Daten des EDA aufgrund von schlechter Qualität in diesem Experiment nicht verwendet werden.

Experiment 2 in Chemnitz: November 2016

Im November 2016 erfolge die Aufnahme der Daten von 68 Probanden (Mittelwert Alter = 22,0; SD = 3,1; 29,4% männlich; 98.6% keine ATC Erfahrung), die an zwei Terminen einer eigens entwickelten Studie teilgenommen haben (siehe Abbildung 25). Das Experiment untersuchte den Einfluss von emotionalen Einflüssen auf die Leistung und Arbeitsbelastung in Fluglotsenaufgaben.

Das 2 (Emotion) x 4 (Ereignisse & Leistung) Innersubjekt-Experimentdesign beinhaltete die Erfassung von Persönlichkeitsmerkmalen (durch BigFive-Fragebogen), die momentane Stimmung der Probanden (durch ASTS-Fragebogen), die aktuell empfundene Arbeitsbelastung (durch ISA-Fragebogen) und die verschiedenen Pupillenerweiterungs- und EDA-Daten aller Probanden.

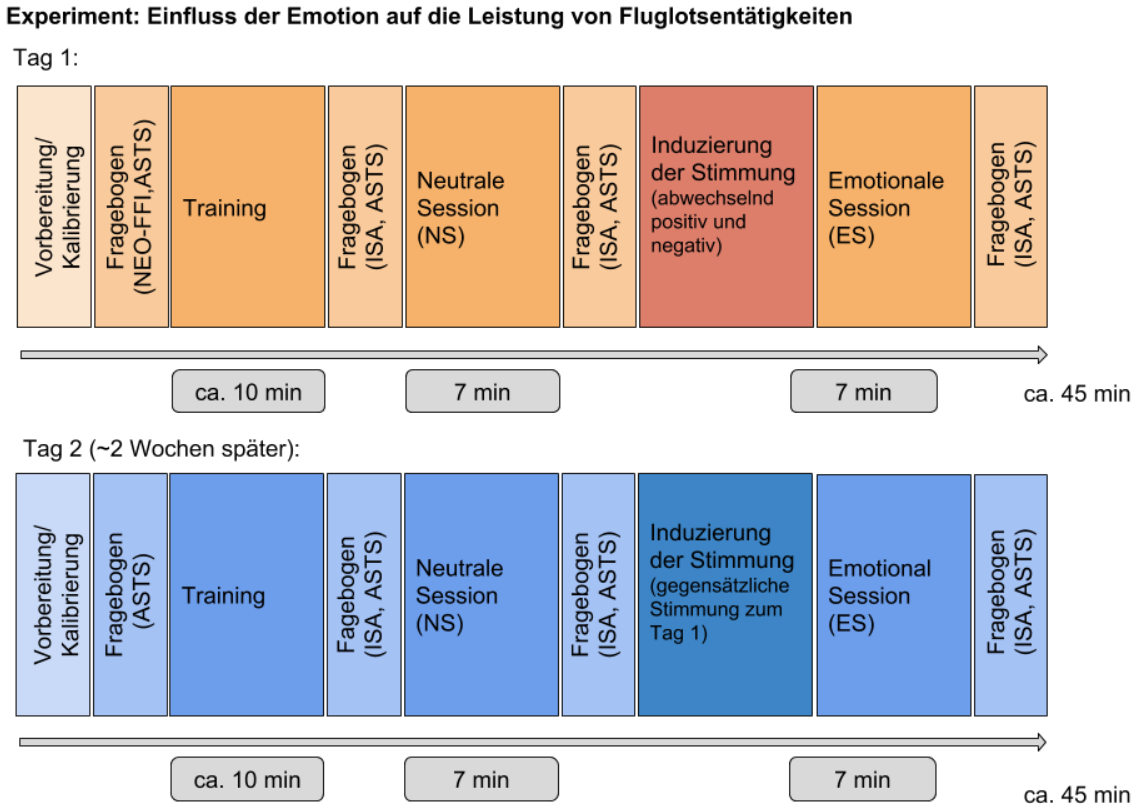


Abbildung 25: Experiment-Prozedur für das zweite Experiment

Experiment 3 in München: November 2017

Im November 2017 erfolgte die Aufnahme der Daten von 20 Fluglotsen (Mittelwert Alter = 35,6; SD = 6,6; 80,0% männlich). Sie durchliefen aus organisatorischen Gründen im Vergleich zu den Studierenden aus Experiment 2 allerdings nur eine Experimental-Session (Abbildung 25). Das Experiment untersuchte, inwiefern die Fluglotsenaufgaben der Simulation mit den Aufgaben der realen Arbeit vergleichbar sind und ob die Fluglotsen die Aufgaben der Simulation aufgrund ihres Expertenwissens besser erledigen können.

Das 1 (Emotion) x 4 (Ereignisse & Leistung) Zwischensubjekt-Experimentdesign beinhaltete die Erfassung von Persönlichkeitsmerkmalen (durch BigFive-Fragebogen), die momentane Stimmung der Probanden (durch ASTS-Fragebogen), die aktuell empfundene Arbeitsbelastung (durch ISA-Fragebogen) und die verschiedenen Pupillenerweiterungs- und EDA-Daten aller Probanden.

Die Pupillenerweiterung wurde in jedem Experiment zum einem als Mittelwert über die gesamte Session gemessen. Zum anderen wird die aufgabenbezogene Veränderungen der Pupille während verschiedener vordefinierter Ereignisse, die während der Simulation auftreten, einzeln berechnet und untersucht. Dabei wurden vier Klassen von Ereignissen definiert. Die erste Ereignisklasse beinhaltet das Erscheinen eines Flugzeugpaars. Es wurde angenommen, dass neue Flugzeugpaare Ressourcen im Arbeitsgedächtnis beanspruchen, woraufhin sich die Komplexität der Aufgabe erhöht und somit auch die Arbeitsbeanspruchung steigt. Die zweite Ereignisklasse beinhaltete Kollisionen, die nicht verhindert werden konnten. Es wurde angenommen, dass die Arbeitsbelastung in diesem Moment steigt, weil Probanden eventuell darüber nachgedenken, welchen Fehler sie gerade begangen haben. Die dritte Klasse von Ereignissen beinhaltete das sichere Verlassen der Flugzeuge vom dargestellten Radar. Bei diesen Ereignissen wurde angenommen, dass die Arbeitslast sich verringert, da die vorher gebundenen Ressourcen im Arbeitsgedächtnis freigegeben werden können und die Aufgabenkomplexität sinkt. Die vierte Klasse beinhaltete die Flugroutenänderungen, die die Probanden mit Hilfe verbaler Kommandos vornehmen und von der Simulation geloggt werden. Es wurde angenommen, dass vor und während einer solchen Aktion die Arbeitsbelastung steigt, da strategisch über mögliche Änderungsoptionen nachgedacht werden muss. Alle diese Ereignisse wurden mit Hilfe der entwickelten Simulationssoftware (siehe Abbildung 26) aufgezeichnet.

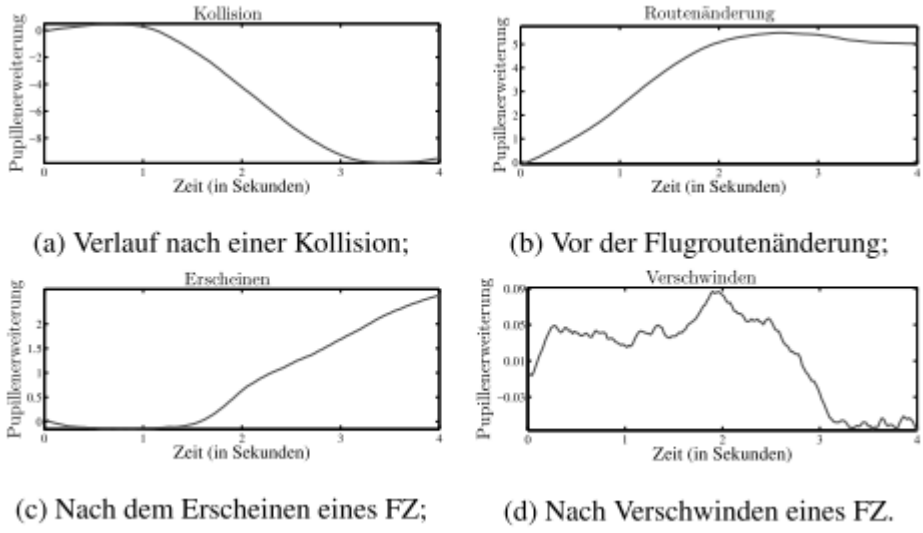


Abbildung 26: Darstellung des mittleren Verlaufs der Pupillenerweiterung über alle Probanden während der verschiedenen Ereignisse.

Die vom Eyetracker aufgezeichneten Datenströme aller Probanden mussten zunächst von Störungen, Zwinkern und anderen ungewollten Artefakten befreit werden, da diese sonst den Mittelwert der jeweiligen Pupillenerweiterung der Sessions verfälschen können. Es wurden dabei Standardmethoden verwendet, welche mit Hilfe von Matlab programmiert wurden.

Nach Aufnahme der Experimente und der visuellen Inspektion der Arousaldaten wurde festgestellt, dass in Experiment 1 viele der Datensätze fehlerhaft waren oder schlicht kein valides Arousalsignal verzeichneten (mehr als 80% der Werte zwischen 0.01 und 0.05). Aus diesem Grund konnte das Arousal in diesem Experiment nicht ausgewertet werden. In dem zweiten Experiment ergab sich bei lediglich 30 Probanden ein valides EDA-Signal. Diese Daten wurden anschließend mit Hilfe von Standardanalysen analysiert, welche LEdaLab – ein Programm für Matlab – bereitstelllt. Dabei wurde das komplexe EDA-Signal in die zu Grunde liegende tonische (skin conductance level: SCL) und in die

schnelle phasische Komponente (SkinConductance Responses: SCRs) zerlegt. Beide Veränderungen der Hautleitfähigkeit resultieren im Allgemeinen aus der sympathischen neuronalen Aktivität. Die SCRs können anschließend gezählt werden und sind wichtige Indikatoren für die emotionale Erregung der Probanden. Im dritten Experiment konnten aufgrund von technischen Problemen auch nur sehr wenige valide Arousaldaten aufgezeichnet werden.

Weil beim ersten Experiment lediglich Pupillen- und Fragebogendaten vorlagen, wurde zunächst ein Modell entwickelt, welche die Verarbeitung von Pupillometrie-Daten und die Erkenntnisse aus Fragebögen integriert. Auf Basis dieser psycho-physiologischen und Fragebogen-Daten wurde mit Hilfe des Modells ein emotionaler und kognitiver Zustand berechnet.

Modellierungsdetails am Beispiel Arbeitsbelastung auf Basis der Pupillendaten

Da das Modell nicht nur zustandsbasierte Aussagen treffen sollte, sondern auch einen dynamischen Verlauf vorhersagen sollte, wurden die Ereignisse in den Pupillometrie- und EDA-Datenströmen auf ihre dynamische Entwicklung untersucht (siehe Abbildung 27).

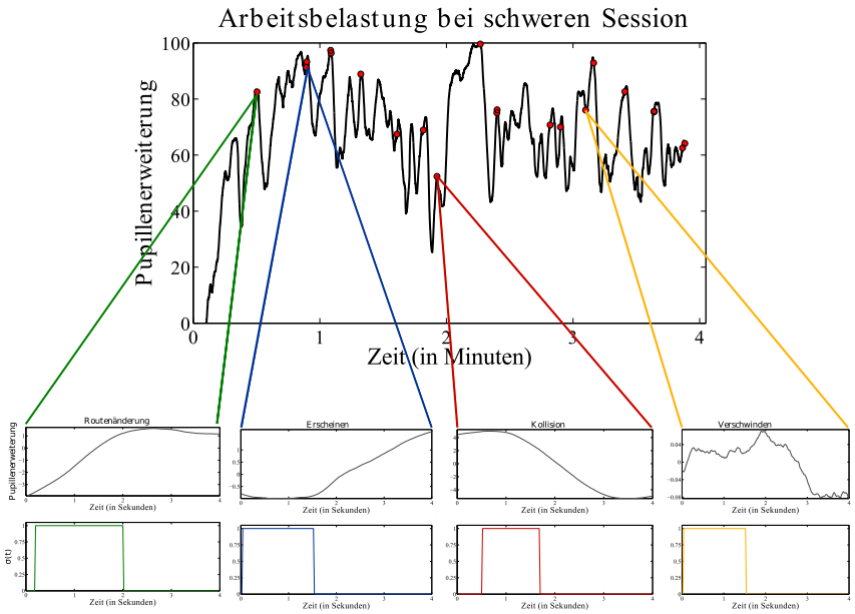


Abbildung 27: Schematische Darstellung anhand welchen Daten das mathematische Modell für die Berechnung der Arbeitsbeanspruchung aufgebaut wurde.

Nach einer Vorverarbeitung wurden die erhobenen Daten auf Merkmalsmuster für die Arbeitsbelastung untersucht. Dafür wurden die Daten in Matlab geladen und mit Hilfe von Zeitreihenanalysen untersucht. Zunächst wurden die Ereignisse zu den aufgezeichneten Zeitpunkten im Datenstrom zugeordnet. Anschließend wurde der Pupillenerweiterungsverlauf während der verschiedenen Ereignisklassen (Verschwinden, Erscheinen, Kollisionen und Aktionen) herauskopiert (Abbildung 29). Jedes so kopierte Zeitfenster umfasste mind. 1,5 – 4 Sekunden.

Anschließend wurde ein dynamisches Modell aufgebaut, welches die Ereignisse als Eingabeströme verwendete und die Pupillengröße als Ausgabefunktion. Mit Hilfe der Matlab System Identification Toolbox wurden zu den unterschiedlichen Eingabe Funktionen entsprechende Transferfunktionen gesucht, die zu der aufgezeichneten Pupillengröße passen (Abbildung 28Abbildung 30).

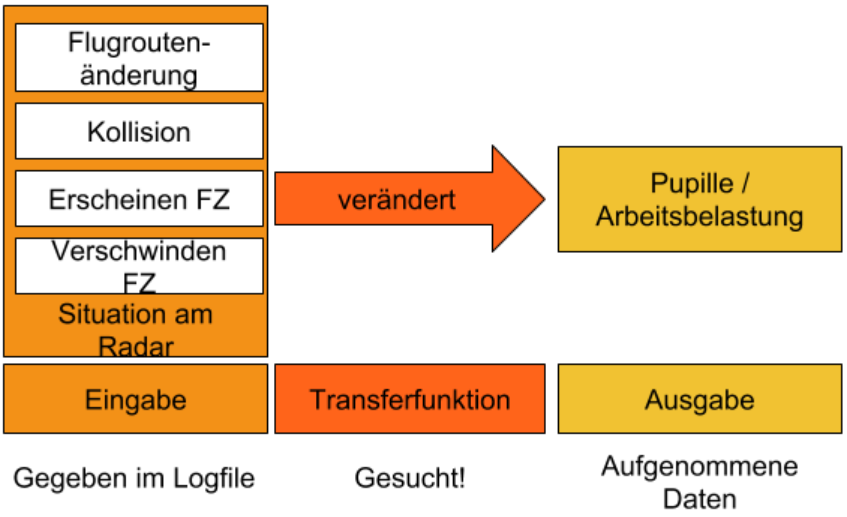
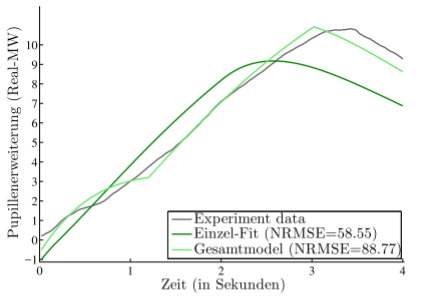
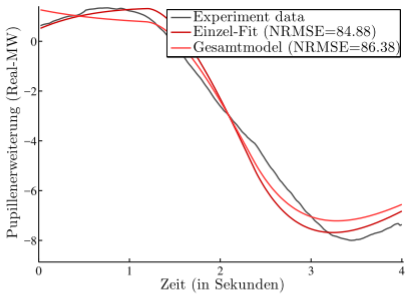


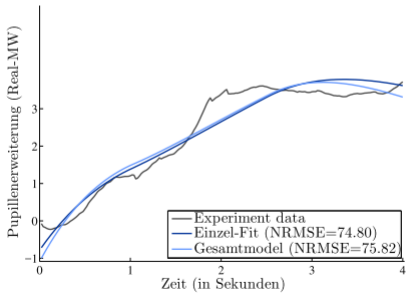
Abbildung 28: Aufbau eines dynamischen Modells mit Hilfe der Matlab-System-Identfication-Toolbox



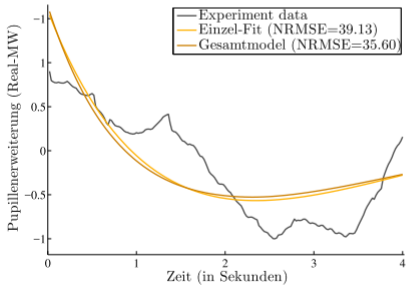
(a) Modell für die Routenänderung.



(b) Modell für Kollisionen.



(c) Modell für Erscheinen von FZ.



(d) Modell für Verschwinden von FZ.

Abbildung 29: Pupillenvergrößerung während verschiedener Events. Die Daten wurden innerhalb eines Simulationsexperimentes mit den Studenten aufgenommen.

Zunächst wurden die Ereignisse und die dazu aufgezeichneten Veränderungen der Pupille einzeln analysiert, um verschiedene Parameter und Zeitkonstanten festzulegen. Die in Abbildung 29 dargestellten Muster wurden zu den verschiedenen Ereignissen detektiert. Die dazu gefundenen Kurvenverläufe, welche aufgrund gefitteter mathematischen Funktionen generiert wurden, sind in Farbe dargestellt. Mit Hilfe dieser entdeckten Muster wurde anschließend ein Gesamtmodell erstellt, welches alle Eingaben verwendet und den gesamten pupillometrischen Ausgabestrom aller Probanden abbilden kann (siehe Abbildung 30).

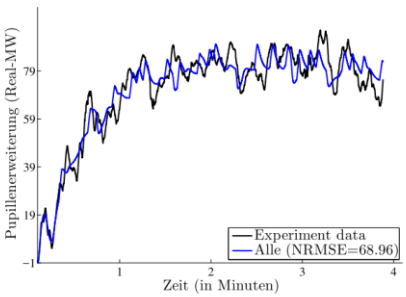


Abbildung 30: Vorhersage der Arbeitsbelastung in der leichten Session von Proband 17.

Das hier beschriebene Modell wurde auf der Konferenz CogSci2017 vorgestellt (siehe auch [Pub18]).

Einbindung aller sensorischen Eingabemodule in das Modell

Die Anpassung und Erweiterung des Modells erfolgte im weiteren Projektverlauf mittels eines iterativen Prozesses, bei dem die in Abbildung 35 dargestellten Kanäle sukzessive implementiert wurden. Um die richtige Einbindung der anderen sensorischen Eingabemodule wie FACS, Sprache und Posen zu realisieren, wurde die Lotsenaufgabe aus Experiment 1 als Grundlage genommen. Dabei dienten den Probanden Videos zur Stimmungsinduzierung: Sie schauten entweder ein lustiges Video (ein Ausschnitt aus „Harry und Sally“) oder ein trauriges Video (ein Ausschnitt aus „Der König der Löwen“). Die Wahl der Videos erfolgte auf der Basis von erfolgreichen Anwendung dieser Ausschnitte in anderen empirischen Studien. Die Wirkung der Videos wurde zudem in einem Pretest nachgewiesen, der online durchgeführt wurde. Die Videos wurden als Manipulationscheck verwendet, um nachzuweisen, dass die einzelnen Kanäle die induzierten Stimmungen auch erkennen können. Außerdem können verschiedene Bias und Gewichtungen für das Modell berechnet werden.

Untersuchung FACS

Die Professur Künstliche Intelligenz lieferte zum zweiten und dritten Experiment eine Analyse des Gesichtes hinsichtlich der fünf Emotionsklassen Freude, Trauer, Überraschung, Ekel und Wut und zusätzlich eine Actionunit-Analyse, welche bestimmte aktivierte Gesichtsmuskeln aufzeichnete. Die gelieferten FACS-Daten wurden analysiert und hinsichtlich der korrekten Erkennung von mehr positiven Emotionen in der positiven Video-Session und von mehr negativen Videos in der negativen Session untersucht (Abbildung 31). Dabei zeigte sich, dass Freude in sehr vielen Fällen erkannt wird. Da die Probanden aufgrund der Arbeitsbelastungsmessung eine Brille trugen wurden auch die einzelnen Actionunits analysiert, um eventuelle Fehldetektionen durch die Verdeckung der Augenpartie auszuschließen.

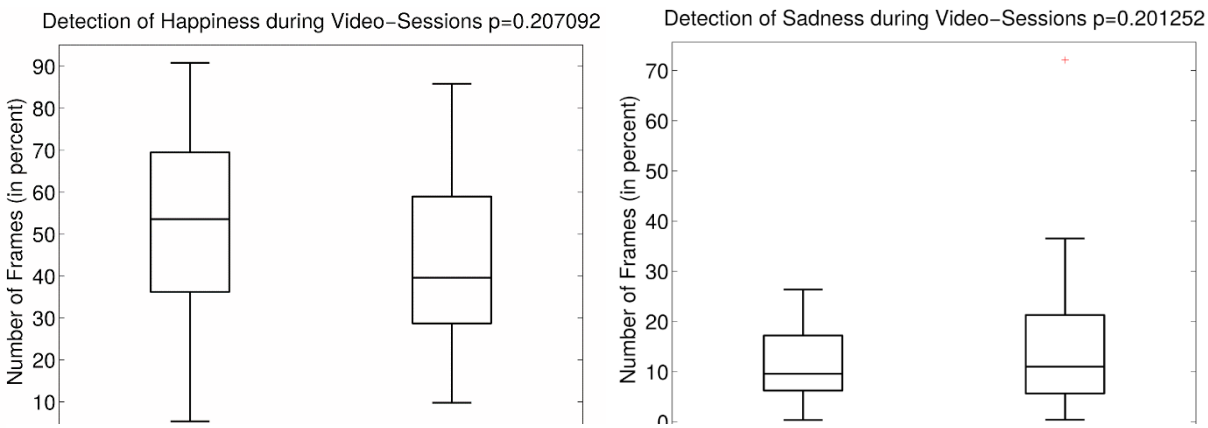


Abbildung 31: Untersuchung der detektierten Emotionen während der stimmungsinduzierenden Experimentierphase.

Diese Analyse zeigte, dass die Mundwinkel-AU immer aktiv war und kein geeigneter Prädiktor für die Stimmung darstellte. Auch die restlichen Actionunits erschienen entweder durch die Perspektive oder die Brille keine genügende Aufklärung zur Stimmungsinduktion zu liefern. Aus diesem Grund konnten die FACS-Daten nur bedingt in das Modell integriert werden.

Untersuchung Posen

Die Professur Graphische Datenverarbeitung lieferte für jeden Zeitpunkt des Experiments verschiedene Posen, die in seltene und häufige (90% im Datenstrom präsent) Posen eingeteilt wurden. Bei der Untersuchung der Posen wurden keine Unterschiede in der Anzahl der aktivierten Posen oder der Anzahl von seltenen Posen in den einzelnen Sessions gefunden. Dies war aber zu erwarten, da Emotionen eher selten in den einzelnen Posen erkannt werden können. Allerdings wurden selten auftretende Posen einzeln untersucht, weil angenommen wurde, dass diese Posen nur in sehr speziellen und alarmierenden Situationen auftreten. Solche Posen könnten einen Indikator für ein steigendes Arousal sein und würden dann in das Modell einbezogen. Dabei wurde die in AP 1.3 (siehe 1. Sachbericht) beschriebene Methode der ereignisbasierten Untersuchung gewählt. In den einzelnen Situationen, wie das Sprechen einer Flugroutenänderung oder die Reaktion nach einer Kollision, ist eine sehr wahrscheinliche Zuordnung zwischen Ursache und Wirkung möglich. Aufgrund der Annahme, dass diese Situationen (z. B. Kommandoerteilung oder ein Kollisionsereignis) spezielle Posen verursachen, wurden die seltenen Posen, die erkannt wurden, während dieser Ereignisse untersucht. Es zeigte sich, dass die Probanden vor dem Geben eines Kommandos eher Denkposen einnehmen (siehe Abbildung 32) und nach einer Kollision, die Hände vor den Oberkörper oder den Kopf halten (siehe Abbildung 33). Im Modell werden aus diesem Grund Posen, die als Denkposen (Finger im Gesicht) klassifiziert wurden, als Indikator für eine erhöhte Arbeitsbelastung gewertet und Posen, die bei Kollisionen gefunden wurden, als Indikator für eine angestiegenes Arousal verwendet.

Abbildung 32: Detektierte Denkposen vor einem gesprochenen Kommando zur Flugroutenänderung.

Abbildung 33: Detektierte Posen, die nach einer Kollision auftraten.

Untersuchung der Audiodaten

Die Professur Medieninformatik liefert zum Zeitpunkt, in welchem ein Kommando gesprochen wurde, einen Aktivitätswert (analog zum Arousal, sowie einen emotionalen Valenzwert, der aussagt, ob der Proband sich eher in einer positiven oder negativen Stimmung befindet. In einem späteren Schritt lieferte die Professur zusätzlich die von der KI auch übermittelten Emotionsklassen (siehe oben). Es wurde die mittlere Valenz und Aktivität in den einzelnen Sessions analysiert. Die Aktivität in der Stimme ist in der negativen Session höher als in der positiven Session (Abbildung 34). In der Valenz wurden keine Unterschiede entdeckt. Hierbei stellt sich aber die Frage, ob die in dem Video induzierte Emotion bei der Aufgabe, die zeitlich später erfolgte, noch erhalten ist. Aus diesem Grund wurden zusätzlich die ersten gesprochenen Kommandos nach dem Video untersucht, denn zu diesen Zeitpunkten müssten die Emotionen noch präsent sein. Die Ergebnisse dieser Analyse waren nicht eindeutig. Dies könnte daran liegen, dass die Emotion in der Stimme generell nicht so stark erkennbar war oder das Modell Defizite hat. Es zeigten sich generell hohe Werte für Überraschung, welche sehr ähnlich Merkmale wie Freude als Entscheidungsgrundlage hat. Um diese Daten allerdings verwenden zu können, müssten deshalb wahrscheinlich die einzelnen Merkmale näher untersucht werden. Für die Modellierung wurden deshalb lediglich der Arousal und Valenzwert in das Modell integriert.

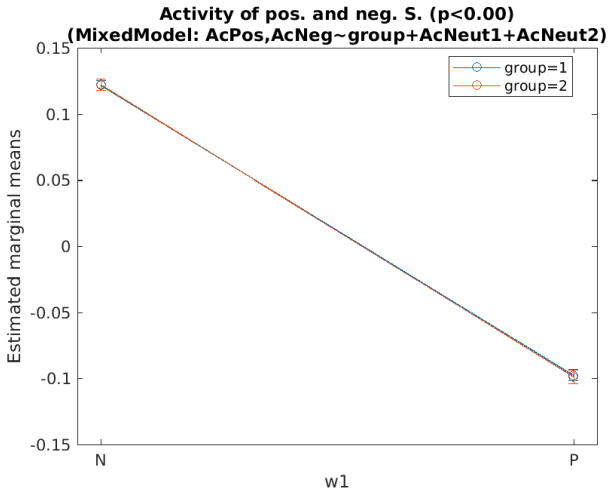


Abbildung 34: Signifikanter Unterschied zwischen negativer Session (N) und positiver Session (P).

Untersuchung der Beziehung der einzelnen Kanäle untereinander

Mittels linearer Regression wurde untersucht, ob FACS und Audio in der Lage sind, den Emotionszustand vorherzusagen, der mit Hilfe von Fragebögen aufgezeichnet wurde. Dabei wurden die letzten Sprachereignisse und die Analyse des emotionalen Ausdrucks in der letzten Minute des Experiments als Grundlage genommen. Es zeigten sich keine signifikanten Ergebnisse, woraus zu schließen ist, dass die Analysen dieser Kanäle nicht die subjektiv empfundene Emotion vorhersagen können.

Des Weiteren wurde untersucht, ob die erkannten Klassen von FACS und Audio miteinander korrelieren. Dabei zeigte sich kein stimmiges Gesamtbild über alle Messzeitpunkte. Audio und FACS erkennen in vielen Fällen unterschiedliche Emotionen. Da schon die Manipulationschecks sehr fragwürdig waren, werden diese Analysen nicht mit im Modell verwendet.

Integration aller Kanäle im finalen Modell

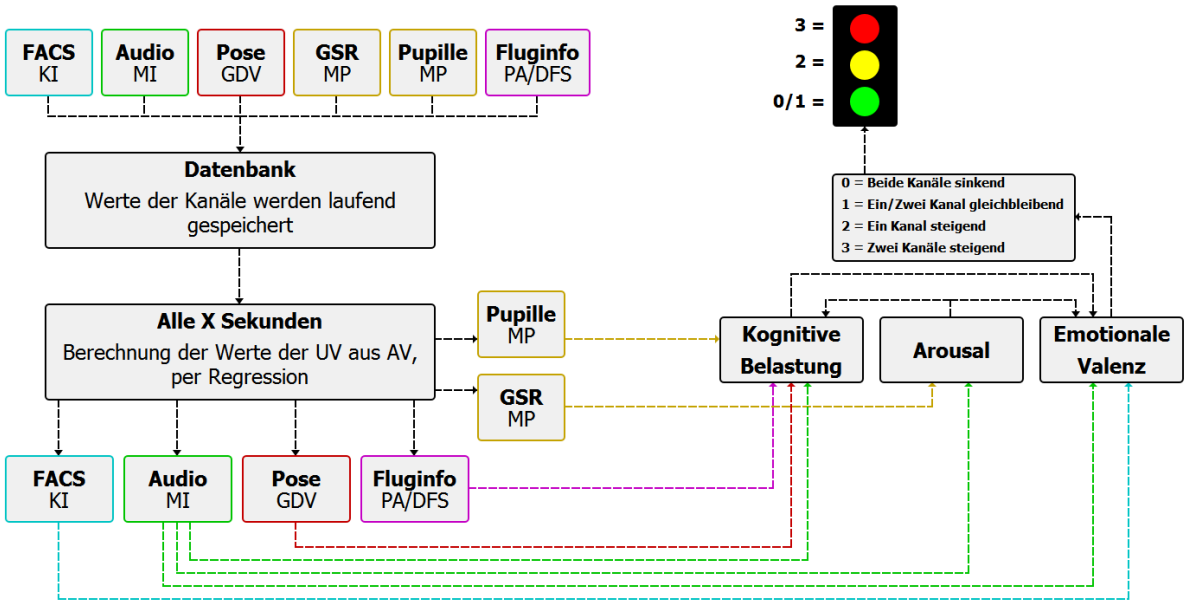


Abbildung 35: Finales Modell der Datenverarbeitung im Prototyp.

Auf Basis der obigen Analysen wurden die Kanäle wie folgt in das Modell (Abbildung 35) integriert:

- Die kognitive Belastung wird als Wert zwischen [0;1] berechnet: sie ergibt sich aus den aufgezeichneten Pupillendaten und einer Erhöhung, wenn seltene Denkpossen erkannt werden.
- Das Arousal wird als Wert zwischen [0;1] berechnet: es wird anhand der EDA-Daten berechnet und beim Auftreten von Kollisionsposen zusätzlich erhöht.
- Die emotionale Valenz wird als Wert zwischen [0;1] berechnet: sie wird durch den Mittelwert zwischen Valenzwert der FACS-Daten (Positiv: detektierte Freude; Negativ: detektierte Wut und Trauer) und der Aktivität der Stimmanalyse berechnet.

Die Sensordatenfusion wurde mittels der vorliegenden Experimentdaten an die Eingaben der Detektionsmodule optimiert. Dafür wurde die Schaltung der Ampel anhand der Fragebögen-Daten angepasst.

Die Ampelfarbe Rot bedeutet, dass ein Ausnahmezustand eingetreten ist und die Supervisoren das Fluglotsen-Team möglichst bald entlasten sollten. Die Ampel wird auf Rot geschalten, wenn:

- Arousal oder Workload > 0,7 oder
- negative Emotion längere Zeit (> 4 min für Exp.) > 0,5,
- Workload längere Zeit (>4 min für Exp.) > 0,5,
- Workload und Arousal > 0,6 sind.

Die Ampelfarbe Gelb bedeutet, dass der Zustand des Fluglotsen nicht kritisch ist, aber bei höherer oder gleichbleibend hoher Belastung kritisch werden kann. Die Ampel wird auf Gelb geschaltet, wenn:

- Arousal oder Workload > 0,5,
- negative Emotion längere Zeit (>2 min für Exp.) > 0,4,
- Workload längere Zeit (>2 min für Exp.) > 0,4,
- Workload und Arousal sind > 0.3.

Die Ampelfarbe Grün bedeutet, dass alles in Ordnung ist und keine kritischen Situationen zu erwarten sind. Die Ampel ist Grün, wenn alle vorher beschriebenen Fälle nicht eingetreten sind.

1.6 Emotions- und Kommunikationsmodellierung (AP 6.5, AP 8)

Zur Vorbereitung der Emotions- und Kommunikationsmodellierung wurden im Laufe des Projekts mehrere Datenerhebungen mit aktiven Center-Lotsen (Standorte: München und Langen) und Mitarbeitern der Abteilung Forschung & Entwicklung der Deutschen Flugsicherung durchgeführt, um einen realistischen Einblick in den Arbeitsablauf von Center-Lotsen zu erhalten. Dabei kam neben Experteninterviews auch die Methode der teilnehmenden Beobachtung im Center-Betrieb zum Einsatz, um Abläufe, Bedingungen, Zusammenarbeit und andere für die Modellierung relevante Aspekte erfassen zu können. Die im Rahmen dieser Datenerhebung gewonnenen Erkenntnisse lassen sich in drei größere Bereiche gliedern: (1) Informationsverarbeitung, (2) Kommunikation und (3) Emotion.

1. Obwohl die Lotsen jeweils die gleichen Arten von Informationen verarbeiten, ergaben sich individuelle Unterschiede bei der mentalen Organisation der Daten. Diese ist vor allem auf die Maximierung von Strukturiertheit und Vorausplanung ausgerichtet, so dass Konflikte durch rechtzeitige Maßnahmen möglichst gar nicht erst auftreten. Dabei haben Center-Lotsen keine 3D-Repräsentation der Flugsituation im Kopf, sondern erkennen auf Basis der Flugpläne, Flugrouten und Berufserfahrung frühzeitig mögliche Konflikte, wobei der 2D-Darstellung der Position der Flugzeuge auf dem Radarschirm eine unterstützende Funktion zukommt. Die individuellen Unterschiede bei der Verarbeitung und internen Repräsentation sind für den Arbeitsalltag insofern irrelevant, als dass alle Center-Lotsen ihre Arbeit zügig, präzise und sicher ausführen. Sie berichteten jedoch von Einflüssen infolge schwankender kognitiver und emotionaler Befindlichkeiten. Diese haben einerseits eine intentional höhere Konzentration und Anstrengung zur Folge, und andererseits wird die Arbeit als anstrengender bis hin zu stressig empfunden. Je nach aktueller Befindlichkeit unterscheidet sich also die subjektive Einschätzung einer objektiv identischen Arbeitslast, was als individueller Faktor im Modell zu berücksichtigen ist, auch wenn eine valide Messung dieser subjektiven und tagesformabhängigen Schwankungen eine große Herausforderung darstellt.

2. In der Arbeit der Center-Lotsen spielt Kommunikation eine wesentliche Rolle. Große Teile der sprachlichen Kommunikation (Face-to-Face, Telefon, Funk), beispielsweise zwischen Executive

und Piloten, sind standardisiert. Dies gilt in eingeschränkter Form auch für den Informationsaustausch zwischen den Lotsen, da auf verbaler Ebene ein definierter Wortschatz existiert, um die für die Arbeit notwendige Spezifik und einen hinreichenden Detailgrad bietet. Intentionale kommunikative Akte finden besonders bei der Abstimmung innerhalb eines Lotsenteams auch mittels funktionaler Gesten statt. Grundsätzlich gibt es bei der tätigkeitsbezogenen Kommunikation infolge der oben genannten Standardisierung ein hohes Maß an Übereinstimmung. Unterschiede sind eher zu erwarten, wenn die Kommunikation keinen direkten Bezug zur beruflichen Tätigkeit hat. Das Ausmaß solcher Gespräche ist von mehreren Variablen abhängig, die intern als Eigenschaften des jeweiligen Lotsen (z. B. Extraversion) bzw. als Eigenschaft der Dyade (z. B. Sympathie für den jeweils anderen Lotsen) oder extern (z. B. Aufmerksamkeits- und Handlungsanspruch der jeweiligen Flugsituation) aufgefasst werden können. Für die Modellierung kann sich auf die tätigkeitsbezogene Kommunikation beschränkt werden, da das entsprechende Verhalten in seiner Art und Häufigkeit in Abhängigkeit objektiver Daten (Flugverkehr im eigenen Sektor und in den Nachbarsektoren) erfassbar und von primärem Interesse ist. Ob und inwiefern die Fluglotsen neben der eigentlichen Tätigkeit kommunizieren, ist nicht relevant und kann zudem nicht modelliert werden, da die dafür benötigten Variablen entweder nicht in ihrer Gänze bekannt oder nicht praktikabel zu erheben sind. Übereinstimmend berichteten Lotsen jedoch davon, dass Episoden privater Kommunikation generell nur in Phasen der Nicht-Auslastung stattfinden, so dass für die Auswertung von Simulationsdaten durch die Dichotomisierung des Kommunikationsverhaltens in tätigkeits- und nicht-tätigkeitsbezogen ein globaler Indikator für das generelle Belastungslevel in einem bestimmten Zeitraum existiert.

3. Wie bereits unter (1) angegeben, kann auch der emotionale Zustand der Center-Lotsen die Leistung beeinträchtigen. Dabei gaben die Lotsen vor allem Stress als Hauptgrund für eine negative Erlebnisqualität an. Als stressauslösend werden dabei vor allem folgende Situationen erlebt:

    1. hohes Verkehrsaufkommen, d. h. eine große Anzahl an Flugzeugen, vertikalen Flugbewegungen und Kursüberschneidungen;

    2. ungeplante Ereignisse, z. B. das Auftauchen eines im Flugplan nicht vermerkten Flugzeugs im Sektor oder die Schließung eines Flughafens;

    3. Ausfall des Equipments, z. B. Ausfall des Radars, des Funks oder der Flugplandaten;

    4. persönliche Gründe, wie zum Beispiel längere Abwesenheit durch Urlaub/Krankheit und die damit verbundene erneute Einarbeitung in das System; Stimmungen und Probleme.

In allen drei Bereichen zeigten sich neben grundsätzlichen Übereinstimmungen zwischen allen Center-Lotsen auch individuelle Unterschiede. Daraus ergab sich die Notwendigkeit, die Modellierung auf zwei

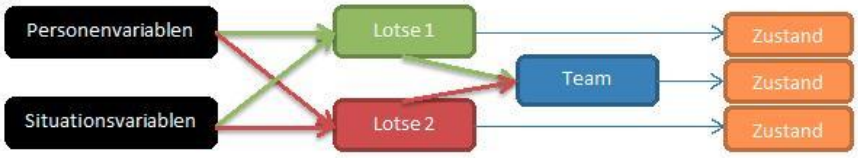


Abbildung 36: Schematische Darstellung Elemente für die Modellierung.

Ebenen anzusetzen: Einerseits musste der kognitiv-emotionale Zustand der individuellen Lotsen modelliert werden, andererseits aber auch der Zustand der Dyade. Dafür wurden die für das Modell benötigten Variablen generell in zwei Gruppen zu unterteilen: Situations- und Personenvariablen. Zur ersten Gruppe gehören objektiv erfassbare Daten wie das Verkehrsaufkommen. Die zweite Gruppe umfasst Konstrukte, die nicht direkt messbar sind, wie zum Beispiel die aktuelle Befindlichkeit der Lotsen. Die Identifizierung der einzelnen relevanten Variablen erfolgte in einer Kombination aus grundsätzlichen psychologischen Überlegungen und der Auswertung der in Simulationen gewonnenen Messdaten. Abbildung 36 zeigt eine vorläufige schematische Darstellung der für die Modellierung relevanten Größen und deren Beziehung zueinander.

Den Ausgangspunkt für die Bildung des Emotions- und Kognitions-Modells einzelner Lotsen bildeten die theoretisch hergeleiteten Variablen, die auch in der Literatur als emotions- oder belastungsbestimmend beschrieben werden (z. B. im vorliegenden Teilmodell Arousal und Pupillengröße). In Abbildung 37 ist ein auf dieser Grundlage entwickelte Modell dargestellt. Anhand der Datenerhebungen wurden diese Variablen auf ihre Richtigkeit hin überprüft und deren Abhängigkeiten bzw. relativer Einfluss auf andere Variablen ermittelt (z. B. den emotionalen und kognitiven Gesamtzustand). Daraus erfolgte eine Gewichtung der Eingabewerte. Außerdem wurden zusätzliche Einflussgrößen hinzugenommen und irrelevante Variablen verworfen. Das so entstandene neue Modell (siehe Abbildung 39) wurde anschließend mit den zusätzlich erhobenen Datensätzen nach demselben Prinzip evaluiert.

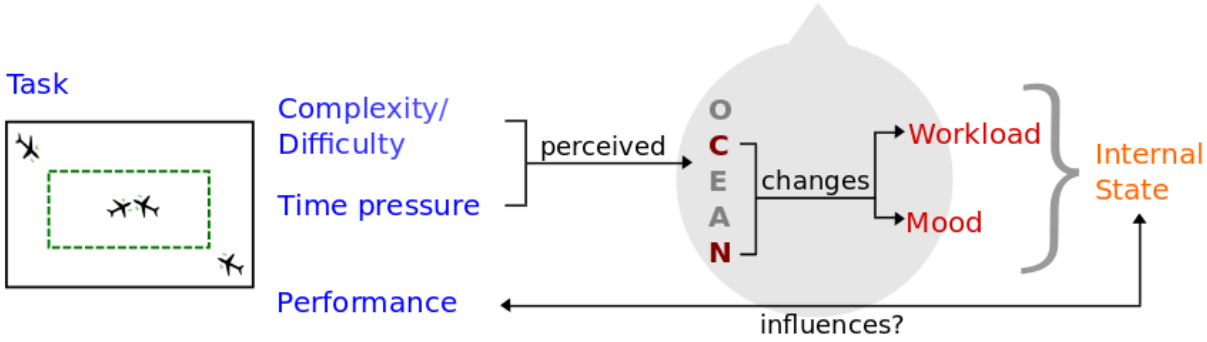


Abbildung 37: Das auf der Literatur basierende Kognitions- und Emotionsmodell

Folgende Erkenntnisse wurden aus den einzelnen Experimenten gewonnen werden:

Experiment 1:

Untersuchung des Zusammenhangs Arbeitsbelastung und Aufgabenschwierigkeit

Es zeigten sich in den unterschiedlichen Schwierigkeitsstufen der Aufgabe unterschiedliche Mittelwerte in der Pupillenerweiterung. So war die Pupillengröße in der schweren Session signifikant größer als in der leichten Aufgabe. Das heißt, dass die Pupillenerweiterung als Indikator für die erhöhte Arbeitsbelastung verwendet werden kann.

Untersuchung des Zusammenhangs Persönlichkeit und emotionale Befindlichkeit

Es zeigte sich, dass vor allem in der schweren Session ein Zusammenhang zwischen Neurotizismus und einer negativeren Empfindungen und einer schlechteren Leistung besteht. Wechselwirkungen zwischen Persönlichkeit und Arbeitsbelastung wurden nicht festgestellt.

Untersuchung des Zusammenhangs Aufgabenschwierigkeit und emotionale Befindlichkeit

Es zeigte sich eine Verschlechterung der Stimmung nach jeder Session [Training (40,6) > LC (39,0) > HC (36,3)]. Abbildung 38 zeigt, dass die Probanden nach jeder Session eine schlechtere Stimmung zeigten (rote Linie) und erschöpfter waren (grüne Linie). Bei der Dimension Ruhig-Nervös (grüne Linie) zeigten sich kaum Veränderungen.

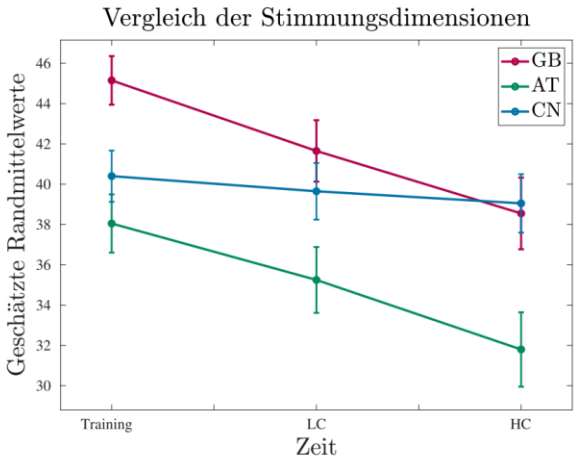


Abbildung 38: Veränderung der emotionalen Befindlichkeit der Probanden während des Trainings, der leichten (LC) und schwierigen (HC) Session (GB: Gut-Schlecht-Dimension, AT: Wach-Müde-Dimension, CN: Ruhig-Nervös-Dimension).

Untersuchung des Zusammenhangs Stimmung auf Arbeitsbelastung

Es wurde kein Zusammenhang gefunden, demzufolge weist die Stimmung in diesem Experiment keine Korrelation zur Arbeitsbelastung auf.

Untersuchung des Zusammenhangs zwischen Arbeitsbelastung/ Stimmung und Leistung

Die Datenanalyse ergab eine statistisch signifikante Korrelation zwischen Leistung und Stimmung: Es zeigte sich, dass eine positiv berichtete Stimmung auch mit einer besseren Leistung einhergeht.

Anhand dieser Erkenntnisse wurde das Modelle entsprechen (Abbildung 39) angepasst:

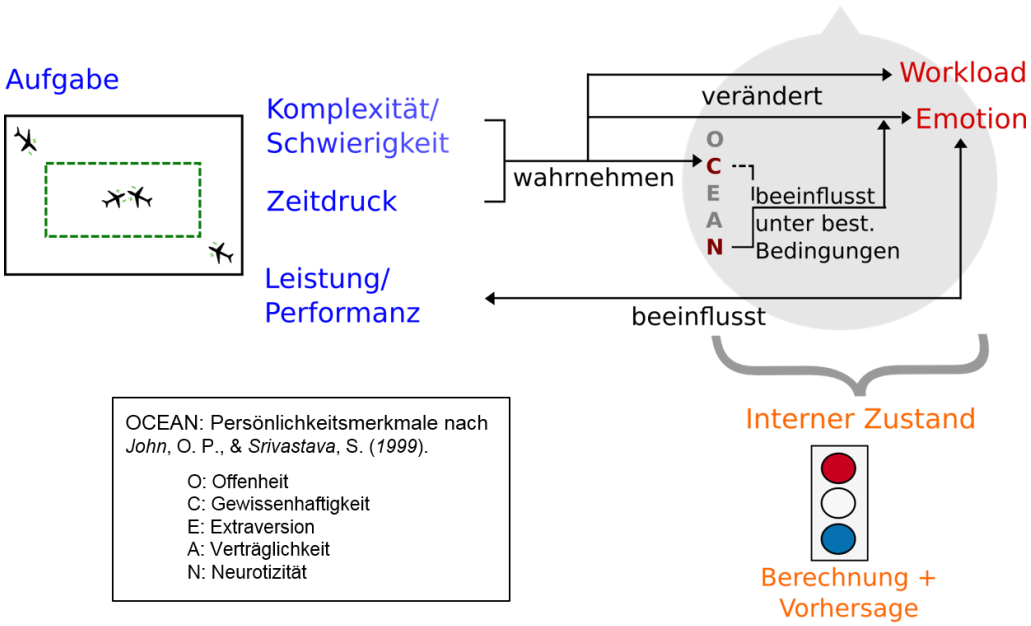


Abbildung 39: Auf Basis von Experiment 1 angepasstes Kognitions- und Emotionsmodell.

Die hier präsentierten Erkenntnisse wurden in einem Journal **Fehler! Verweisquelle konnte nicht gefunden werden.**veröffentlicht.

Experiment 2

Untersuchung Zusammenhang Stimmung und Leistung

Negative Emotionen akquirieren zusätzliche analytische Ressourcen, welche zu einer verbesserten Performanz führen. Die empfundene Emotion korreliert nur bei neuen Aufgaben mit positiveren Emotionen.

Die nachfolgend präsentierten Erkenntnisse wurden auf der Konferenz HCII2017 präsentiert [Pub20].

Die induzierte Stimmung hat in der ersten Experiment-Session eine größere Auswirkung als in der zweiten Session. Dies bedeutet, dass Training bzw. Übung den Einfluss von Emotionen verringern können. Die Stimmung am Ende der Session wird von der Leistung und von der Persönlichkeit beeinflusst. Die Probanden fühlten sich besser, wenn sie eine gute Leistung erbracht haben und schlechter, wenn sie eine schlechte Leistung erbracht haben. Dieser Einfluss ist allerdings – je nach Persönlichkeitsausprägung in Neurotizismus und Gewissenhaftigkeit – unterschiedlich. Die objektive Arbeitsbeanspruchung ist unabhängig von der Stimmung. Dies bestätigte die Ergebnisse aus Experiment 1. Allerdings ist die subjektiv empfundene Arbeitsbelastung (ISA) maßgeblich von der Persönlichkeit und der aktuellen Stimmung abhängig.

Es erfolgte anschließend die Integration aller Experimentdaten aus den durchgeführten Experimenten. Dazu mussten die aufgezeichneten Daten in die Prozessmodeldarstellung überführt werden. Dazu wurden für die verschiedenen Ereignisse (Erscheinen, Verschwinden, Routenänderung und Kollision) Eingabeströme erstellt, die die geloggten Ereignisse in einer Step-Funktion darstellen (siehe Abbildung 40). Das unter „Datenfusion“ beschriebene Modell wurde zur Vorhersage der Arbeitsbelastungen aller Probanden genutzt.

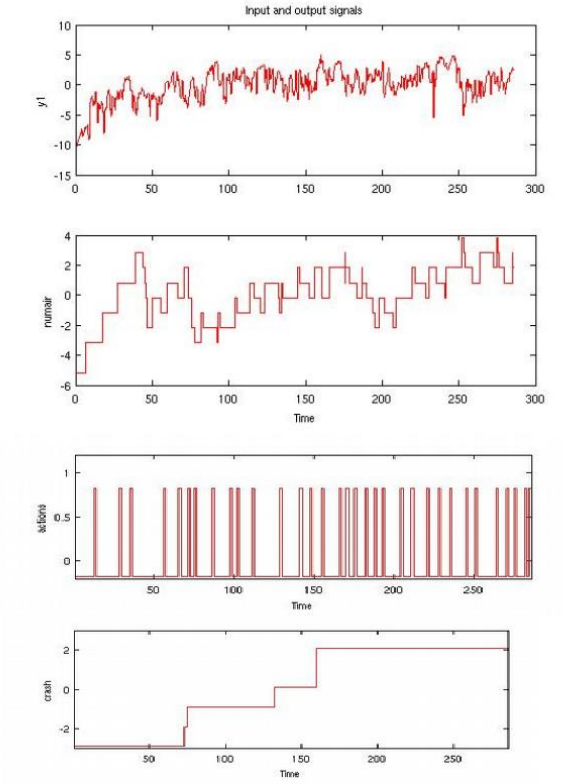


Abbildung 40: Darstellung der modellierten Eingabe- und Ausgabefunktionen des dynamischen Prozessmodels. Diese Informationen dienten als Grundlage, um die Parameter des dynamischen Modells zu bestimmen, d.h. um zu berechnen, wie ein Crash oder ein gesprochenes Kommando die Pupillengröße verändert bzw. den Workload erhöht.

Modellierungsdetails für die Stimmung

Aufgrund der statistischen Auswertungen aus den Experimenten können folgende Schlüsse für die Berechnung der Stimmung anhand der einzelnen Faktoren gezogen werden: Aufgabenschwierigkeit, Stimmung zu Beginn der Sessions, Arbeitsbelastung/Diameter, Anzahl der Kollisionen, Neurotizismus und Gewissenhaftigkeit haben einen Einfluss auf die verschiedenen Dimensionen der Stimmung.

Mit Hilfe von Regressionsgleichungen wurden diese Einflüsse und deren Gewichtung berechnet und für die Modellierung der Stimmung verwendet.

Für die Vorhersage der Stimmung in der Gut-Schlecht-Dimension konnte eine Regressionsgleichung gefunden werden ($F(1,31) = 7,36$; $p < 0,001$), die mit einem $R^2 = .57$ rund 57% der Varianz in dieser Dimension vorhersagen kann. Die vorhergesagten Gewichte der Probanden betragen dabei:

$50.60 - 9.99(\text{Schwierigkeitsgrad}) - 0.43(\text{Pupillengröße}) + 0.72(\text{GBstart}) - 1.19(\text{Anzahl der Kollisionen}) - 3.55(\text{CWert}) - 2.24(\text{NWert}) + 0.26(\text{Schwierigkeitsgrad} * \text{Pupillengröße}) + 0.23(\text{Anz. Kollisionen} * \text{CWert})$.

Die Gleichung wurde durch schrittweises Hinzufügen und Entfernen der einzelnen Terme gefunden.

Für die Vorhersage in der Stimmung in der Wach-Müde-Dimension wurde folgende Regressionsgleichung gefunden:

$-73.9 - 0.617(\text{PupillengrößeBeginn}) + 1.46(\text{PupillengrößeEnde}) + 2.98(\text{ATstart}) - 1.589(\text{NWert}) - 0.037(\text{ATstart} * \text{PupillengrößeEnde}) + 0.015(\text{ATstart} * \text{Diastart})$.

Diese Regressionsgleichung ($F(1,33) = 17,3$; $p < 0,001$) zeigte mit $R^2 = .715$ eine Varianzaufklärung 75%.

Die Stimmungsdimension Ruhig-Nervös wurde lediglich von dem Ruhig-Nervositätswert vor der Session beeinflusst, und deswegen wurde diese Dimension nicht modelliert.

Die Regressionsgleichungen in der Wach-Müde und Gut-Schlecht-Dimensionen waren von der objektiven Arbeitsbelastung und der Anzahl der Kollisionen während den Sessions abhängig. Diese Werte wurden kontinuierlich über die gesamten Sessions aufgezeichnet. Auf der Basis dieser Daten konnte nun auch die Stimmung in einem kontinuierlichen Verlauf abgebildet werden. Dazu wurden alle 5 Sekunden ein Mittelwert über die Anzahl der Kollisionen und die Pupillenerweiterung gebildet. Anschließend wurde mit der Regressionsgleichung für diese Punkte die vorhergesagte Stimmung berechnet. Ein Beispiel ist in Abbildung 41 dargestellt. Mit Hilfe dieses Verfahrens konnte aus einer Zustandsmessung eine kontinuierliche Messung/ Simulation gewonnen werden.

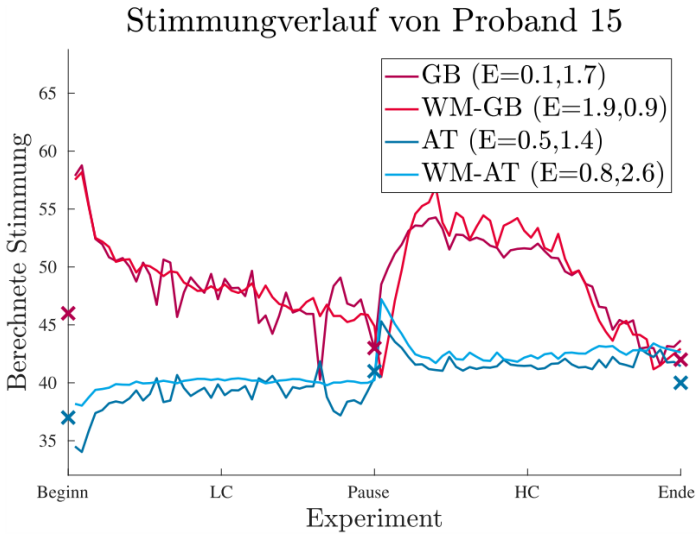


Abbildung 41: Berechnung des Stimmungsverlaufs während der leichten und schweren Session anhand der im Experiment gefundenen Regressionsgleichungen.

Das **kognitive Kommunikationsmodell** bildet die relevanten Interaktionen zwischen der Lotsendyade ab und erlaubt in der Kombination mit den Daten über das emotional-kognitive Befinden der einzelnen Lotsen und externen Parametern eine Bewertung des Zustands der Dyade. Als Ausgangspunkt für das kognitive Kommunikationsmodell dient das sogenannte *mutual belief model* für kooperatives Arbeiten (Abbildung 42), welches für das Centerlotsen-Szenario als gut geeignet gilt [149].

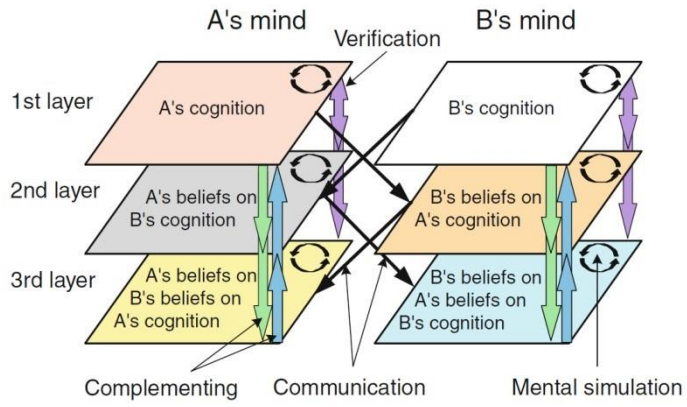


Abbildung 42: Mutual belief model für kooperatives Arbeiten nach [149].

Abbildung 42 verdeutlicht, dass sich dieses Modell neben den jeweils eigenen Kognitionen der Personen A und B auf den Ausschnitt der mentalen Ebene konzentriert, in dem die auf den Kooperationspartner bezogenen Kognitionen und Überzeugungen angesiedelt sind. Daneben können aber auch Interaktionen bzw. Kommunikation in der Lotsendyade durch den emotionalen Zustand und durch die entsprechenden Überzeugungen zum emotionalen Zustand des Partners in analoger Weise zu den Kognitionen ausgelöst werden. Darüber hinaus können aufgabenspezifische Kognitionen sowie deren Einflussfaktoren das Kommunikations- und Interaktionsverhalten induzieren bzw. moderieren, so dass diese ebenfalls für das Modell von Relevanz sind.

Die konkreten Formen der Kommunikation betreffend, so lassen sich diese sinnvollerweise in Übereinstimmung mit der Literatur (z. B. [110]) wie folgt unterscheiden: Frage, Antwort, Aussage, Anweisung, Rückmeldung. Diese Formen können in den Ausdrucksarten verbal, nonverbal oder in Kombination von beiden Ausdrucksarten auftreten. Im weiteren Projektverlauf wurde auf Basis des Modells untersucht, ob bzw. inwiefern es systematische Zusammenhänge zwischen den Inhalten, Formen und Ausdrucksarten der Kommunikation gibt und wie sich diese gegebenenfalls in Abhängigkeit der kognitiven und emotionalen Zustände der Lotsen verändern und dieses sich in den einzelnen Kanälen manifestiert.

1.7 Evaluierung und Anpassung (AP 16.3, AP 18)

Vorhersagegüte des Modells

Das in AP 8.2 entwickelte Modell wurde anhand der aufgenommenen Daten evaluiert. Zuerst wurde getestet, wie viele Eingabeströme relevant sind für das Ermitteln eines bestmöglichen Workload-Modells (siehe Abbildung 43). Es zeigte sich, dass jenes Modell die besten Ergebnisse erzielt, welches alle Eingabeströme (Anzahl der Flugzeuge auf dem Radar, Erscheinen und Verschwinden der Flugzeuge, Crashs und Aktionen) berücksichtigt (grüne Linie).

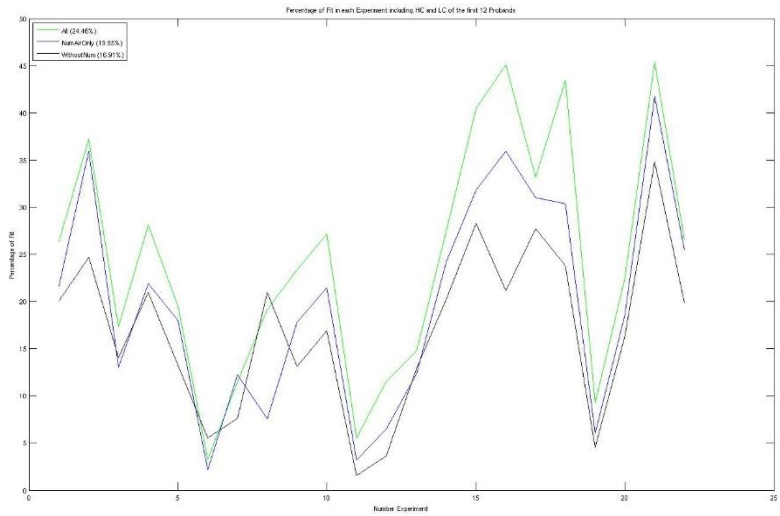
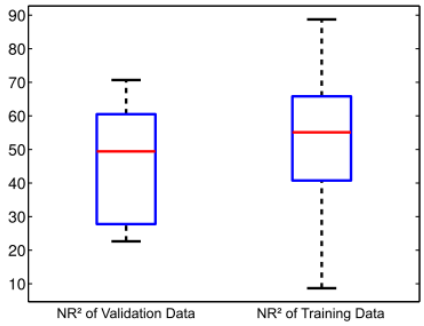


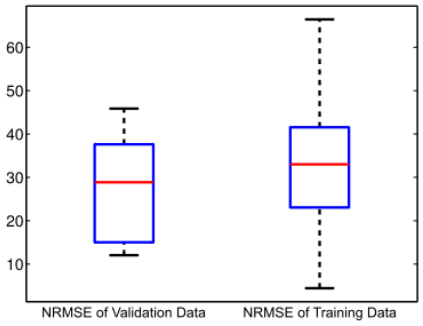
Abbildung 43: Untersuchung, welche Daten für ein geeignetes Modell benötigt werden. Die schwarze Linie stellt die Ergebnisse des Modells dar, welches die Eingabeströme Crash, Aktion sowie das Erscheinen und das Verschwinden der Flugzeuge

enthält. Die blaue Linie enthält die Ergebnisse des Modells, welche nur die Anzahl der Flugzeuge auf dem Radar enthält. Die grüne Linie zeigt die Ergebnisse des Modells, die alle Eingabeströme beachten.

Anschließend wurde ermittelt, wie gut das entwickelte generelle Workload-Modell den Workload bislang nicht im Modell verwendeter Probanden vorhersagen kann. Dabei wurde der erhobene Datensatz in einen Modellierungs- und Validierungsdatensatz unterteilt. Anschließend wurde untersucht, wie gut das auf dem Trainingsdatensatz ermittelte Modell den Workload der Validierungsdatensätze beschreiben kann (Abbildung 44). Um beide Datensätze miteinander zu vergleichen, wurde von allen Probanden zum einen der normierte mittlere quadratische Fehler (NRMSE) zwischen Workload des Modells und der Pupillenveränderung berechnet und zum anderen das Bestimmtheitsmaß (NR^2) bestimmt. In Abbildung 44 werden die verschiedenen goodness-of-fit-ergebnisse des Modellierungs- und des Validierungsdatensatze dargestellt. Es zeigt sich, dass der normierte, mittlere quadratische Fehler im Mittel bei 28.87% und bei dem Validierungsdatensatz bei 32.99% lag. Bei dem Blick auf den NR^2 zeigt sich, dass das Modell 55.10% der Varianz in der Pupillenerweiterung im Modellierungsdatensatzerklären kann und im Validierungsdatensatz 49.42%.



(a) Comparison of NR^2.



(b) Comparison of *NRMSE*.

Abbildung 44: Vergleich der Modellgüte zwischen Modell und Validierungsdatensatz. a) Vergleich der Varianzaufklärung. b) Vergleich des normierten mittleren quadratischen Fehlers

Das Modell zeigte nach der Anpassung an die neueren Experimente somit gute Ergebnisse bei der Vorhersage neuer Probanden. Der Rest der Varianz wird durch andere Variablen erklärt, vor allem im Bereich der Persönlichkeits- und Situationsvariablen, welche aber im Laufe der Untersuchungen nicht erhoben werden konnten (z. B. Tagesform, Aufregung, Motivation etc.). Im Folgenden wird genauer auf die Validierungen der einzelnen Modell-Komponenten bzw. der anderen Probandengruppen eingegangen.

Validierung des Workload-Modells anhand der Daten im zweiten Experiment

In Abbildung 45 sind die aufgezeichneten mittleren Pupillenveränderungen und die Vorhersage dieser Veränderung durch das Barcelona-Modell dargestellt. Es wurde für die Evaluation des auf dem ersten Experiment basierenden Modells die neutrale Session des neuen Experimentes verwendet, da diese am ehesten den Bedingungen des besagten Experiments entsprechen. Auf dem ersten Blick ist zu sehen, dass die Pupillenerweiterung überschätzt wird. Das heißt, dass Modell nimmt eine größere Pupillenerweiterung an, als es tatsächlich gibt. Die Ursachen sind wahrscheinlich die veränderten Lichtverhältnisse sowie die unterschiedlichen Wertebereiche und Messungsunterschiede bei den Eyetrackern.

Es zeigt sich allerdings auch, dass sich die in Experiment 1 gefundenen Muster auch im zweiten Experiment wiederfinden. So steigt die Pupillengröße bei Ereignissen wie dem Erscheinen der Flugzeuge und bei einer Routenänderung an und sinkt bei der Kollision oder dem Verschwinden von Flugzeugen.

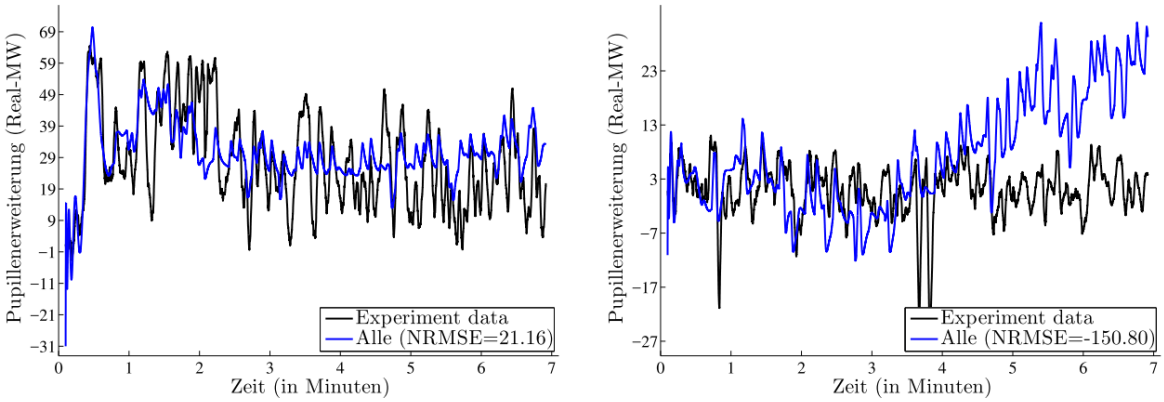
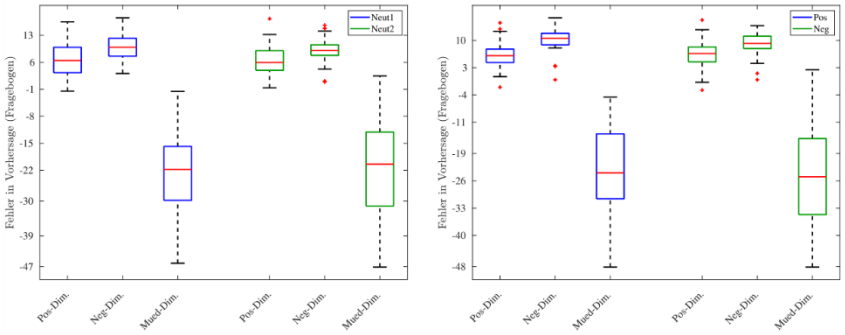


Abbildung 45: Simulation der Arbeitsbelastung in der ersten neutralen Session des 2. Experiments.

Validierung des Stimmungs-Modells anhand der Daten im zweiten Experiment

Hier ergaben sich einige Veränderungen zum ersten Experiment, da ein anderer Fragebogen verwendet wurde. In diesem Fragebogen sind die positive und negative Stimmungsdimension getrennt, es gibt keine Ruhig-/ Nervös-Dimension, da sich diese Dimension als nicht relevant erwiesen hat. Die Wach-/ Müde-Dimension ist mit der Erkennung der Müdigkeitsskala vergleichbar. Die Daten wurden für die Vorhersage mit der Regressionsgleichung aus dem Barcelona-Experiment wie folgt aufgearbeitet: Für die Startparameter in den jeweiligen Stimmungen wurden die Fragebogenergebnisse des vorherigen Zeitpunktes verwendet. Die positive und negative Stimmung wurde mit der Gut-/ Schlecht-Regressionsgleichung vorhergesagt. Dabei wurden je nach induzierter Stimmung entweder nur die positiven Stimmungsfragen verwendet und nur die negativen. Die Müdigkeit wurde mit der Wach-/Müde-Regressionsgleichung vorhergesagt, lediglich diesmal auf Basis der Müdigkeitsdimension des ASTS-Fragebogens. Die Pupillendaten haben durch die Normalisierung einen ähnlichen Wertebereich wie im ersten Experiment, so mussten diese nicht weiterbearbeitet werden. In Abbildung 46 sind die

Fehler in der Vorhersage der verschiedenen Stimmungsdimensionen in den verschiedenen Sessions im zweiten Experiment dargestellt.



(a) Fehlerquote in den neutralen Sessions.

(b) Fehlerquote in den emotionalen Sessions.

Abbildung 46: Fehlerquote bei der Vorhersage der Stimmung in den verschiedenen Sessions des zweiten Experimentes. Die Vorhersage wurde mit Hilfe der in Experiment 1 gefundenen Regressionsgleichung berechnet.

Es zeigt sich, dass die positive Stimmung in allen Sessions relativ gut vorhergesagt werden konnte. Die Müdigkeit in dem Experiment zeigt eine größere Fehlerstreuung. Und in der negativen Dimensionen zeigen sich vor allem in große Streuungen in den Stimmungs-induzierten Sessions sowie der zweiten neutralen Session. Die große Fehlerstreuung in der Müdigkeitsvorhersage können entweder an den Unterschieden in den Fragebögen oder auch daran liegen, dass das Experiment in der virtuellen Umgebung aufgrund der veränderten Übersichtlichkeit und Darstellung der Informationen ermüdender als in einem Computerlabor ist.

Anpassung des Stimmungs-Modells anhand der Daten im zweiten Experiment

Die im ersten Experiment gefundenen Regressionsgleichungen wurden anhand der neuen Daten angepasst und anschließend die Prognose dieses Modells für die einzelnen Probanden berechnet. In Abbildung 47 ist exemplarisch der vorhergesagte Stimmungsverlauf von Proband 34 dargestellt. Es zeigt, sich, dass mit Hilfe der neuen Regressionsgleichung die Stimmung in den einzelnen Dimensionen sehr viel besser vorhergesagt werden konnte.

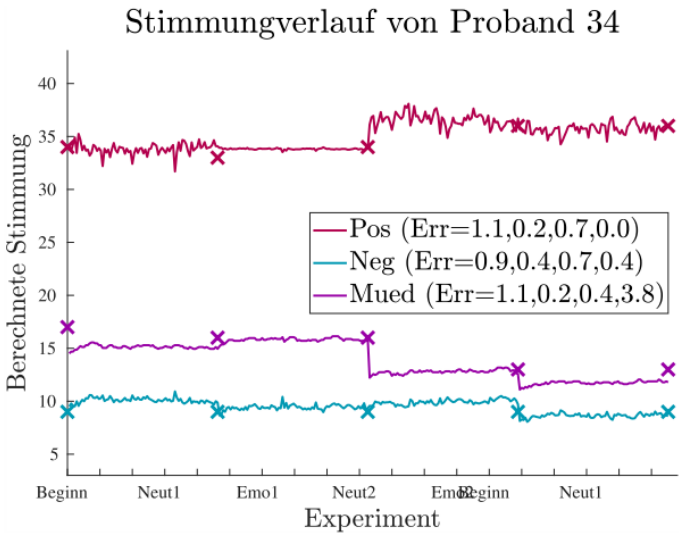


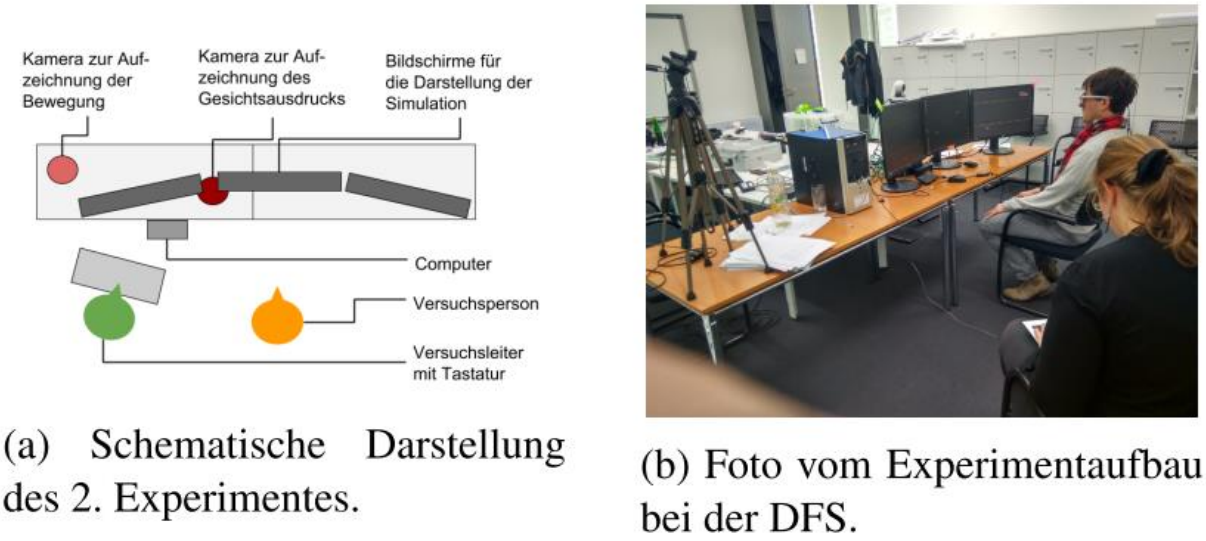
Abbildung 47: Exemplarische Darstellung der Stimmungsverläufe für alle Sessions.

Evaluation des Stimmungs- und Workloadmodells anhand der Expertengruppe (Fluglotsen)

Diese evaluative Auswertung der Daten von Experiment 3 erfolgte in der Verlängerung der Projektphase von Februar bis April 2018.

Experimentelles Vorgehen

Jeder Teilnehmer der Studie war aktiver Fluglotse der DFS vom Standort München. Alle Teilnehmer nahmen freiwillig an der Studie teil und wurden von ihrem Arbeitgeber für dieses Experiment freigestellt. Auch die Räumlichkeiten stellte die DFS innerhalb des Fluglotsencenters zur Verfügung. In einem Besprechungsraum wurden somit Monitore, Rechner und alle Aufzeichnungsgeräte installiert, so dass für die Fluglotsen sehr kurze Wege von der Arbeit zum Experiment bestanden. Im Experimentraum angekommen musste jeder Fluglotse zunächst eine Einverständniserklärung ausfüllen und bekam eine kleine Einführung in das Testsetting. Anschließend erfolgte das Ausfüllen der ersten Fragebögen (1.ASTS, Allgemeine Daten und der Persönlichkeitsfragebogen). Danach erfolgte die Anbringung der verschiedenen Messinstrumente. Nachdem alle Geräte auf korrekte Aufnahme hin überprüft wurden, begann das Experiment mit der Übungssession, bei welcher die Fluglotsen lernten, mit der Simulation, den reduzierten Flugroutenänderungen und der vereinfachten Aufgabe umzugehen. Nachdem die Fluglotsen sich mit der Materie vertraut gemacht hatten, folgten – wie in Experiment 2 – immer abwechselnd Fragebögen und Simulationsaufgabe. Nachdem neutrale und emotionale Session beendet waren und die verschiedenen Messinstrumente entfernt wurden, wurden die Fluglotsen gebeten, einen weiteren Fragebogen auszufüllen, welcher sowohl Gemeinsamkeiten als auch Unterschiede von Simulation und realer Fluglotsenarbeit vergleichen sollte. Danach erfolgte eine kurze Verabschiedung.



(a) Schematische Darstellung des 2. Experimentes.

(b) Foto vom Experimentaufbau bei der DFS.

Abbildung 48: Versuchsaufbau des Experiments bei der DFS.

Untersuchung Vergleich Studenten und Fluglotsen:

Fluglotsen zeigen eine signifikant höhere Ausprägung der Persönlichkeitseigenschaft Gewissenhaftigkeit. Sie zeigen gering höhere Werte im Neurotizismus. Die Studenten erreichten in unserer Fluglotsensimulation rund 100 Punkte mehr als die Studenten – demzufolge scheint die Simulation schon gewisse Prozesse der Fluglotsenarbeit abzubilden. Sie zeigen geringere Veränderung in der positiven und negativen Emotion. Wobei die negativen Emotionen sehr viel geringer sind als bei den Studenten, obwohl der subjektive empfundene Stress bei den Fluglotsen signifikant höher. Dies spricht dafür, dass die Fluglotsen während ihrer Ausbildung lernen vor allem negative Emotionen zu kontrollieren und eventuell zu unterdrücken. Die erhöhte subjektive Belastung kann mehrere Gründe haben. Zum einen die sehr stark vereinfachte Darstellung ihrer Arbeit, was dazu führt, dass sie ihren Handlungsspielraum sehr stark einschränken müssen. Zum anderen waren die Fluglotsen sehr stark irritiert von den Kollisionen – die für sie eine sehr viel stärkere Assoziation von Fehlern, die in ihrer realen Arbeit verlorenen Menschenleben bedeuten. Die hier entdeckten Änderungen wurden in das Fluglotsenmodell integriert.

Untersuchung Vergleichbarkeit Simulation und reale Fluglotsenarbeit

Die Simulation wurde im Allgemeinen von allen Fluglotsen als belastender, schwieriger, stressiger und anstrengender als ihre reale Arbeit empfunden, wobei sich hier kleine Unterschiede zwischen den Fluglotsen ergaben (siehe Abbildung 49). So schätzten Fluglotsen auf der Approach-Position näher an ihrer Arbeitsaufgabe ein als Fluglotsen im Center, welche die höheren Lufträume betreuten. Dies liegt wahrscheinlich daran, dass das Experiment sehr reaktiv ist. Die Bewegungen in hohen Lufträume sind sehr gut planbar und haben selten reaktive Komponenten. Dies ist auf der Approach-Position anders. Dort werden die Flugzeuge einzeln nacheinander auf die entsprechenden Höhen und Linien gebracht, damit sie nacheinander landen können, dieses sogenannte „auffädeln“ beinhaltet das verarbeiten und managen von Flugzeugen auf einem engen Raum, was die Arbeit wiederum reaktiver macht. Deswegen ist die Simulation näher an diesem Arbeitsfall. Des Weiteren wurde gefragt inwieweit die Aspekte Arbeitsbelastung, Anstrengung, Stress, Schwierigkeitsgrad und Zeitdruck der Fluglotsentätigkeit in der Simulation abgebildet wurde. Es zeigt sich, dass die Fluglotsen die einzelnen Aspekte eher mittelmäßig

abgebildet fanden. Vor allem die Geschwindigkeit und die sich daraus ergebende reaktive Komponente kritisierten sie stark, genauso die Geräuschkulisse der kollidierenden Flugzeuge.

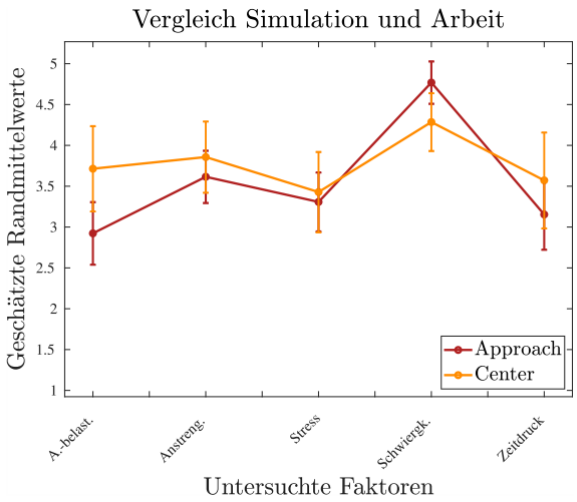


Abbildung 49: Einschätzungen über die Schwierigkeit der Simulation im Vergleich mit der realen Aufgabe (1- Simulation ist viel einfach; 5- Simulation ist viel schwieriger).

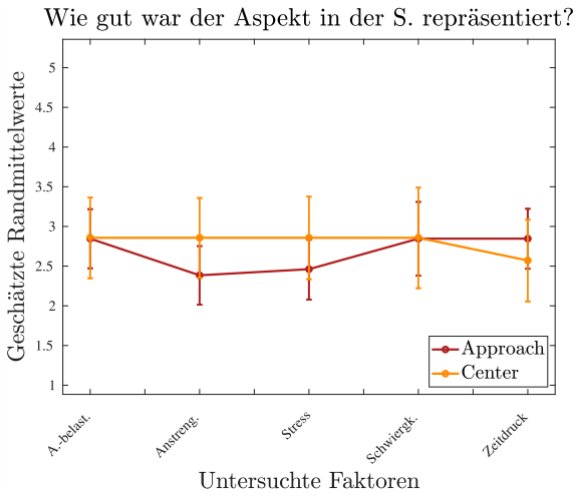


Abbildung 50: Einschätzungen über die Vergleichbarkeit der Simulation mit der realen Fluglotsenarbeit (1- sehr schlecht repräsentiert; 5 – sehr gut repräsentiert).

Ursprünglich war vorgesehen, das Gesamtsystem in Zusammenarbeit mit der DFS im Experimentalbetrieb bei Lotsenplanspielen zu evaluieren. Jedoch stellte sich heraus, dass die Durchführung eigener Experimente an den Systemen der DFS aus technischen, rechtlichen und betriebsinternen Gründen nicht möglich war. Diese Einschränkung hat aber keinen Einfluss auf das Methodeninventar als solches, sondern stellt lediglich eine Einschränkung für die Validierung für das Fluglotsen-Szenario dar.

Zur Evaluierung des Kommunikations- und Interaktionsmodells wurde parallel zu Experiment 3 am DFS-Standort München eine offene, nicht-teilnehmende Beobachtung der Centerlotsen im laufenden Betrieb durchgeführt. Die Notwendigkeit dafür ergab sich einerseits daraus, dass in den Simulationen die Position des „Planners“ nicht hinreichend realistisch nachgebildet erschien. Andererseits erachteten wir eine erneute Datenerhebung als sinnvoll, da seit den ersten lotsenbezogenen Datenerhebun-

gen zu Projektbeginn im Laufe der Modellentwicklung neue Erkenntnisse zu relevanten Variablen gewonnen wurden, die zu diesem Zeitpunkt berücksichtigt werden konnten. Dabei wurden innerhalb von drei Tagen insgesamt sieben einstündige Beobachtungen mit folgenden Foki vorgenommen: dreimal „Planner“, einmal „Executive“ und dreimal „Executive“ und „Planner“. Dabei wurden in den Einzelsitzungen folgende Daten erhoben: Proaktive Interaktionen mit dem Partner (dienstlich und privat), Telefongespräche, Absprachen mit anderen Lotsen (am eigenen Platz sowie anderen) und längere Abwesenheiten vom Arbeitsplatz. Während sich auf der Position des „Executives“ zwischen den Simulationen und Realbetrieb bei Art und Anzahl der Tätigkeiten zumindest hinreichende Parallelen fanden, zeigten sich bei der Position des „Planners“ große Abweichungen:

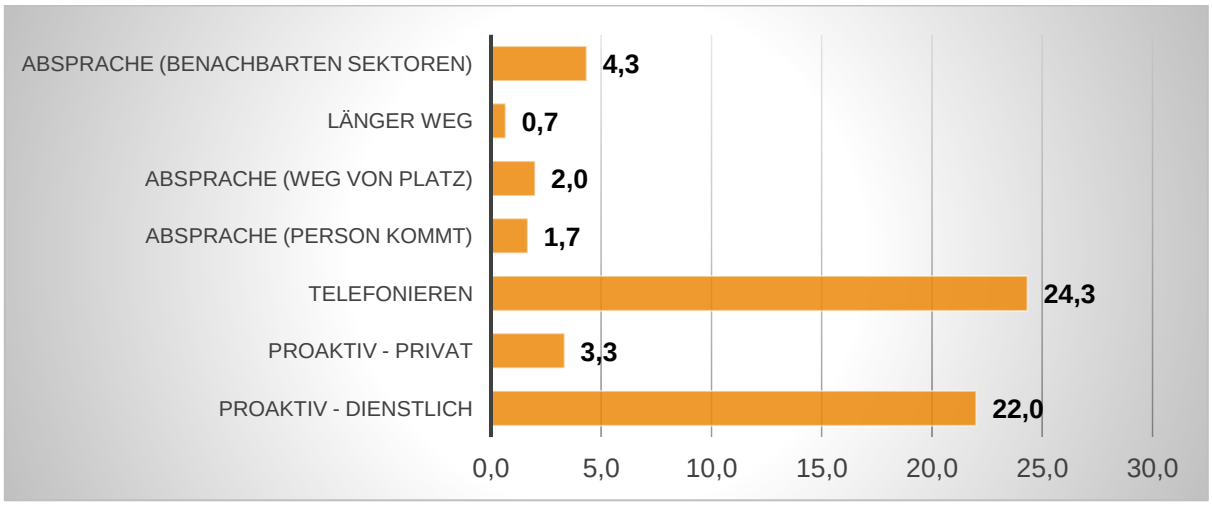


Abbildung 51: Durchschnittliche Häufigkeit kommunikativer Akte des „Planners“ pro Stunde.

Deutliche Unterschiede des Realbetriebs zu den Simulationen bestehen vor allem darin, dass die häufigste Tätigkeit, das Telefonieren zur Abstimmung mit anderen Sektoren, in den Simulationen komplett fehlt. Gleiches gilt für alle anderen Arten von Absprachen sowie die Häufigkeit der Kommunikation innerhalb der Lotsendyade insgesamt (siehe Abbildung 51). Noch deutlicher tritt der Unterschied zutage, wenn man bedenkt, dass sich etwa ein Drittel der kommunikativen Akte in den Simulationen auf die Funktionalität des Systems beziehen und nichts mit der eigentlichen Lotsenarbeit zu tun haben.

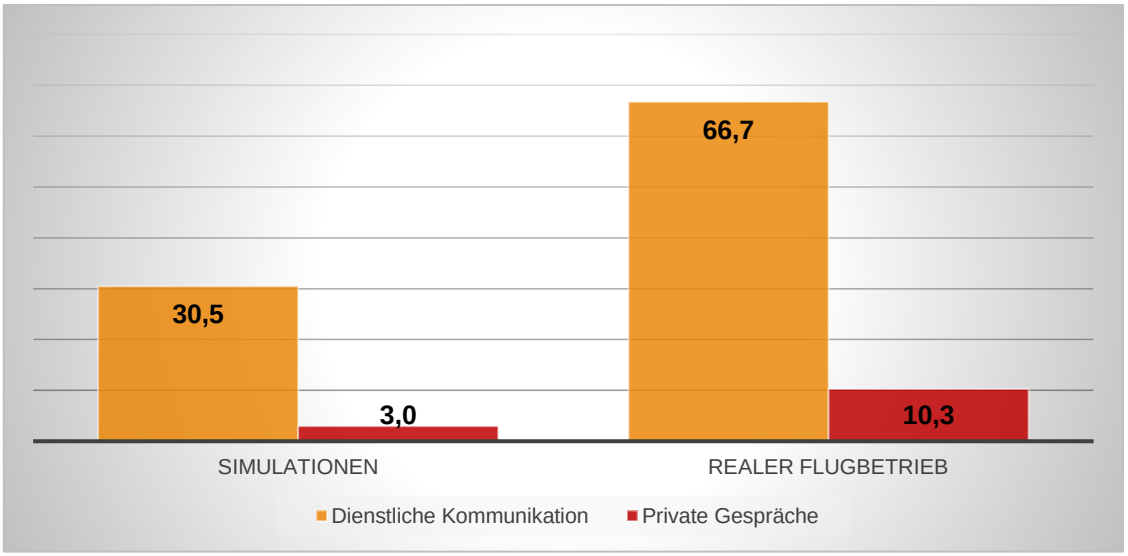


Abbildung 52: Vergleich der Kommunikationsarten und –häufigkeiten zwischen „Executive“ and „Planner“ bei Simulationen und im realen Flugbetrieb.

Als wichtiges Ergebnis innerhalb des Projekts muss also festgehalten werden, dass eine extern valide Simulation des Kommunikations- und Interaktionsverhaltens nicht möglich ist. Um dem allgemeinen Interesse der möglichst optimalen Arbeitsbedingungen von Centerlotsen gerecht zu werden, sollten für zukünftige Studien in diesem Bereich die Möglichkeiten von Datenerhebungen während des realen Flugbetriebs erwogen werden, da die daraus gewonnenen Erkenntnisse und Verbesserungsmöglichkeiten im Vergleich zu arbeitsrechtlichen und die Sicherheit betreffenden Vorbehalten womöglich überwiegen. Generell ist die Kommunikation und Interaktion zwischen den Centerlotsen ein essenzieller Bestandteil der Arbeitsabläufe, der basierend auf jahrelanger Erfahrung und Optimierung im Praxisbetrieb größtenteils einwandfrei und effektiv funktioniert. Jedoch ergab unsere Beobachtung vor Ort, dass es bei der Kommunikation zwischen den Lotsen benachbarter Sektoren noch Verbesserungsbedarf und –möglichkeiten gibt: Generell gilt die Vorgabe, dass alle Kommunikationsprozesse zwischen den Lotsen aufgezeichnet werden müssen, damit bei Vorfällen eine genaue Untersuchung und Rekonstruierbarkeit gewährleistet ist. Dies wird aber gerade bei kurzen Absprachen zwischen benachbarten Sektoren nicht immer eingehalten, da es offenbar einfacher ist, direkt miteinander zu sprechen als mit einer direkt benachbarten Person zu telefonieren. Hier kollidieren offensichtlich die vorgegebenen Regeln mit der Praktizierbarkeit, so dass diesbezüglich Verbesserungspotenzial besteht. Obwohl diese Aktivitäten außerhalb des Fokus des Projekts liegen, möchten wir dieses für den Kontext der Flugsicherung relevante Ergebnis nicht vorenthalten.

Implementierung von neuen Erkenntnissen in das Modell und Nutzerakzeptanz

Wie bereits im vorherigen Abschnitt beschrieben, ergab sich durch die Einschränkungen bei der Durchführung einer szenarionahen Evaluation der Bedarf neuerlicher Datenerhebungen. Diese wurden nach abgeschlossener Auswertung in einem weiteren Schritt zur Aktualisierung bzw. Optimierung des Modells verwendet (siehe oben). Des Weiteren wurden im Verlaufe des Projekts diverse Variablen identifiziert, die für das Interaktions- und Kommunikationsverhalten und damit auch das Modell relevant sind, uns aber nicht zugänglich waren. Dazu gehören insbesondere Persönlichkeitsvariablen der Lotsen, wobei sowohl stabile Dispositionen als auch dynamische Merkmale (Tagesform, Eingewöhnungsbedarf nach Urlaub oder Krankheit etc.) relevant sind. Des Weiteren müssen bei der dyadischen Arbeit generell auch teambezogene Variablen berücksichtigt werden, etwa die Sympathie für den Teampartner und die Arbeitserfahrung als Team, da diese das Kommunikations- und Interaktionsverhalten

ebenfalls beeinflussen können. Derlei Daten liegen potenziellen Anwendern bezüglich des eigenen Personals im besten Fall vor bzw. können mit Standardinstrumenten erhoben werden, so dass die moderierende Wirkung im praktischen Einsatz berücksichtigt werden kann. In Ermangelung der Verfügbarkeit aller relevanten Daten werden diese Variablen nicht in das finale Modell integriert, da eine reine Auflistung ohne weiterführende Untersuchung keine konkreten und praktisch brauchbaren Aussagen zulässt. Ihre moderierende Wirkung kann von Fall zu Fall unterschiedlich groß sein, so dass hier eine globale Aussage bezüglich ihrer Wichtigkeit nicht pauschal erfolgen kann. Stattdessen muss bei der Anwendung des Modells jeweils für den konkreten Fall abgewogen, welche personen- bzw. teambezogenen Variablen erhoben und berücksichtigt werden können und sollten. Deren Erhebung und Auswertung gehört zu den Standardmethoden der Psychologie, unterliegt wie in unserem Projekt aber oft den oben genannten rechtlichen oder betrieblichen Einschränkungen.

Bezogen auf die Evaluierung der Nutzerakzeptanz des Systems, die insbesondere im Rahmen der ELSI-Workshops in Bremen (2016, projektintern) und Regensburg (2017, im Rahmen der Mensch & Computer) thematisiert wurden, müssen zwei Aspekte unterschieden werden: (1) die Zustandsmessungen an sich und (2) die Verwendung der erhobenen Daten.

(1) Bezogen auf die eigentlichen Messungen gibt es seitens der Nutzer ein generelles Einverständnis für ein solches System, da es das Ziel der Verbesserung der Arbeitsbedingungen dient. Allerdings wird seitens der Arbeitnehmer verlangt, dass die Messgeräte möglichst minimal invasiv sind und die Arbeit in keiner Weise behindern dürfen. Kritisch wurde dabei vor allem der Einsatz der Blickbewegungsbrille gesehen, da diese das natürliche Blickverhalten und damit den für die Informationsaufnahme essenziellen visuellen Kanal behindert. Die für die Messung physiologischer Parameter existierenden Lösungen per Armband wurden als akzeptabel bewertet, wobei jedoch die Qualität und Zuverlässigkeit der Daten hier im Verhältnis zum Komfort abgewogen werden muss. Zukünftige Generationen von Wearables scheinen eine insgesamt bessere Qualität zu bieten und dieses Problem zu lösen. Die für die anderen Kanäle notwendigen Messinstrumente (Kameras und Mikrofon) wurden per se als nicht invasiv akzeptiert.
(2) Die Verwendung der Daten wird von Seiten der Arbeitnehmer als problematisch eingestuft. Generell wird hier das Potenzial des Datenmissbrauchs bzw. die Verwendung der Daten für andere Zwecke als so hoch eingeschätzt, dass der Betriebsrat dem Einsatz eines solchen Assistenzsystems in der Praxis nicht zustimmen würde. So ist beispielsweise eines der Bedenken, dass die durch das System erhobenen Daten seitens der Betriebsleitung dazu verwendet werden könnten, die Performance der Mitarbeiter zu messen und ggf. bei Gehaltsverhandlungen oder anderen Personalentscheidungen heranzuziehen. Die Fragen, ob diese Bedenken für alle potenziellen Anwenderkreise solcher Assistenzsysteme generalisierbar sind und wie dieses Problem generell gelöst werden kann, sind nicht Gegenstand des Projekts und müssen demnach an anderer Stelle beantwortet werden.

1.8 Informationsvisualisierung und Mensch-Maschine-Interaktion (AP 7, AP 12, AP 17)

In dem Projekt wurden zwei Ansätze verfolgt, um die ermittelten Emotions- und Kommunikationszustände der Fluglotsen für deren Arbeit unterstützend einzusetzen:

a) Eine Rückkopplung der Daten in die Benutzeroberfläche der Fluglotsen in Form einer adaptiven Nutzerschnittstelle und
b) eine Visualisierung der Systemausgaben für die Wachleiter (Supervisoren) der Fluglotsen, welche weitere unterstützende Maßnahmen einleiten können.

Im Folgenden werden die entwickelten Prototypen, die zugrundeliegenden Konzepte und die verwendete Methodik beleuchtet.

Analyse der aktuellen Nutzerschnittstelle für Fluglotsen

Die eingehende Untersuchung der Arbeitsweise der Fluglotsen und der bestehenden Nutzerschnittstellen, sowie die Ermittlung ihrer Stärken und Schwachstellen bilden die Grundlage für die Entwicklung einer intuitiven adaptiven Nutzerschnittstelle. Wichtige Aspekte der Untersuchungen im Hinblick auf die Emotions- und Arbeitslastadaptivität waren insbesondere die Ermittlung stress- und emotionsintensiver Situationen, den damit verbundenen Gefahren und der in der Praxis gängigen Verfahren, mit diesen Situationen umzugehen.

Die Daten zur aktuellen Arbeitssituation wurden während folgender Aktivitäten erfasst:

- Beobachtungen und Interviews mit Fluglotsen im DFS Center Langen (Februar 2015)
- Beobachtungen und Interviews mit Fluglotsenschülern der DFS Akademie in Langen (Februar 2015)
- Beobachtungen und Interviews im DFS Center München (April 2015)
- Fokusgruppe/ Workshop mit Fluglotsen in München (April 2015)
- Dokumentationsanalyse der Dokumentation des Münchener Systems (Juni 2015)
- Beobachtungen und Interviews im DFS Center Karlsruhe (August 2015)
- Videoanalyse von Simulationsläufen mit Fluglotsen in der Forschungsabteilung der DFS (Juni 2015, September 2015, November 2015)

Die gesammelten Daten wurden kodiert und kategorisiert. Dabei folgten wir aufgrund des explorativen Ansatzes keinem standardisierten Kodierungsschema, sondern kodierten die Daten nach einem induktiven Ansatz mit Hilfe eines Affinitätsdiagramms.

In Bezug auf Emotions- und Stressdaten konnte zunächst festgehalten werden, dass die Fluglotsen an einer direkten übersichtlichen Darstellung der aktuellen Arbeitslast der umliegenden Sektoren sowie der eigenen Arbeitslast interessiert sind. Die Information zur Arbeitslast der anderen Sektoren wird als hilfreich eingeschätzt, da sie einen Hinweis liefert, wann Sonderanfragen an die Kollegen angemessen sind oder um die Kollegen notfalls durch die Umleitung von Flugzeugen durch andere Sektoren zu entlasten. Die Information des eigenen Stresswertes kann wiederum eine Hilfe beim Treffen der Entscheidung nach Abkürzungen für einzelne Flieger oder der Annahme anderer Sonderanfragen sein.

Während unserer Untersuchungen wurde mehrfach Unmut über die Erreichbarkeit von Informationen und Funktionalitäten und insbesondere der allgemeinen Übersichtlichkeit der aktuellen Nutzerschnittstellen geäußert, was eine generelle Überarbeitung der Darstellungsweise und Interaktionen nahelegt.

Während der Beobachtungen fielen insbesondere immer wieder Probleme mit der Verwendung der Distanzmessungstools auf. Des Weiteren wurde häufig ein überflüssiges Zoomen und Scrollen der Karten, um weiter entfernte Flugzeuge zu finden, beobachtet.

Die wichtigsten Informationen für die Fluglotsentätigkeit sind die folgenden flugzeugbezogenen Daten in dieser Reihenfolge: 1.) das Rufzeichen, 2.) die dreidimensionale Position, 3.) das Heading (Flugrichtung), 4.) die Steigrate und 5.) die Geschwindigkeit. Allerdings impliziert die zugrundeliegende dreidimensionale Flugsituation nicht automatisch die Bevorzugung einer dreidimensionalen Anzeige der Positionen der Flugzeuge, da unsere Untersuchungen zeigten, dass die mentalen Modelle der Fluglotsen nicht unbedingt dreidimensional sind. Unsere Lotsen beschrieben ihre innere Repräsentation der Flugsituation vielmehr als eine zweidimensionale Situation, welche mit einer zusätzlichen Variable für die Höhe der Flugzeuge angereichert ist (2,5 Dimensionen). Auch in [182], [183] und [187] wurde festgestellt, dass jeder Lotse beziehungsweise Lotsenschüler seine individuelle Repräsentation der Flugsituation aufbaut. Und [186] weisen darauf hin, dass vielmehr die Zugänglichkeit der benötigten Informationen, wie beispielsweise der Höhe, eine Rolle spielt, als die Wahl der Dimensionen der Darstellung.

Weiterführende Informationen zur Auswertung der Beobachtungen und Interviews wurden im Projektverlauf in [114] und [115] publiziert.

Um die bereits qualitativ gesammelten Erkenntnisse auch in Zahlen greifbar zu machen wurde eine Onlineumfrage unter den Fluglotsen durchgeführt, in welcher sie die Wichtigkeit von Informationen im Bezug zu ihrem Stresslevel bewerteten. 19 Fluglosten (16 männlich und 3 weiblich) bewerteten 52 Informationselemente auf einer Fünf-Punkt-Likert-Skala in zwei Konditionen. Im Schnitt waren die Teilnehmer 38,5 Jahre alt und hatten 14,4 Jahre Berufserfahrung mit einer Standardabweichung von je 8,5. Die deskriptive Statistik weist darauf hin, dass die bereits aus den Interviews und Beobachtungen benannten flugzeugbezogenen Daten von höchster Wichtigkeit in Hochlast- wie auch in den Standardsituationen sind.

Die Bewertungen wurden auf Unterschiede der Wichtigkeit zwischen einer Hochlast- und einer Standardsituation untersucht. Da die Annahmen für parametrische Tests nicht erfüllt worden sind, wurde auf non-parametrische Tests zurückgegriffen. Der Wilcoxon-Test zeigte Variablen, die als signifikant weniger wichtig in stressigen Situationen empfunden werden, als in Standardsituationen. Dabei wurde jeweils eine mittlere bis hohe Effektgröße festgestellt. Zunächst fiel auf, dass Daten, die hauptsächlich einer Serviceleistung gegenüber Kollegen oder Piloten dienen, in Stresssituationen weniger wichtig sind. Dies sind beispielsweise die kommende Sektorenfolge (3→2, p=0,0003, r=-0,582), der Zielflughafen (4→2, p=0,0026, r =-0,4879), die vom Piloten gewünschte Flughöhe (4→3, p=0,0086, r=-0,4260) oder der Stresslevel der Kollegen (4→2, 0,0034, r = -0,4758)). Auch werden koordinative Informationselemente, wie Einflugkoordinationspunkte (3→2, p=0,0122, r=-0,4068) oder die aktuelle Uhrzeit (4→1, p=0,0004, r=0,5775) weniger wichtig in Stresssituationen. Diese beiden Datentypen (Serviceinformationen und koordinative Informationen) in Kombination mit der Abnahme der Wichtigkeit von initialen Konfliktindikatoren wie der geplanten Überflugzeit (3→2, p=0,0017, r=-0,5078) und der Information über relevante Intersections (4→2, p=0,0025, r=-0,4899) können auf eine veränderte Arbeitsweise von vorausplanend nach reaktiv hindeuten, bis die Fluglotsen irgendwann ihr mentales „Picture“ vollständig verlieren. Letzterer Zustand wird generell als bedenklich gesehen [3]. Auch werden die Flugobjekte mit steigendem Stresslevel anscheinend immer mehr als gleichartig wahrgenommen. So verlieren der Flugzeugtyp (4→3, p=0,0167, r=-0,3881), der Transpondercode (4→2, p=0,0009, r=-0,5388), die Wirbelschleppenkategorie (4→3, p=0,0028, r=-0,4854), die Art eines Fluges (3→2,

p=0,0051, r=-0,4541), dessen Flugregeln (4→1, p=0,00683, r=-0,4388) und der Flugplanstatus (4→1, p=0,0008, r=-0,54508) stark an Wichtigkeit. Die durch die Onlineumfrage nachgewiesenen Veränderungen in der Wichtigkeit der unterschiedlichen Informationen für die Arbeit der Fluglotsen sind ein möglicher Ansatzpunkt für die adaptive Unterstützung durch die Benutzerschnittstelle.

Die Entwicklung von Bedien- und Visualisierungskonzepten

Die Analyse der Denkprozesse der Lotsen und ihrer Arbeitsweise an den aktuellen Nutzerschnittstellen lieferten mehrere Ansätze um die Lotsenarbeit zu unterstützen. Diese Unterstützung umfasst zum einen eine Überarbeitung verschiedener Designaspekte des bestehenden Interfaces um die Belastung der Lotsen (durch beispielsweise unübersichtliche Interfaces) von vorne herein gering zu halten, zum anderen sollen die Lotsen während fordernden Situationen bestmögliche Unterstützung durch ihre Kollegen sowie durch eine adaptive Anpassung in der Darstellung ihrer Interfaces erhalten. Von den verschiedenen Aspekten, welche wir in den Vorstudien identifizierten (Übersicht in Tabelle 10), konzentrierten wir uns insbesondere auf die Unterstützung während einer fordernden Situation und ein allgemeines gut strukturiertes Arbeitsplatzdesign. Diese Aspekte wurden iterativ untersucht und Lösungsvorschläge dazu entwickelt

Vorbeugung belastender Situationen	Unterstützung während einer fordernden Situation
• schnellere Verfügbarkeit von Höheninformationen • Darstellung von geplanten Flugobjekten, die sich noch außerhalb des aktuellen Radarbildes befinden • allgemeines Arbeitsplatzdesign • Entfernungseinschätzung auf dem Radar • übersichtliche Wetterdarstellung	• Informationsvisualisierung über die Belastung der umliegenden Sektoren • Anpassen der Streifeninformationen (flugzeugbezogene Daten) bei Überforderung • Zusätzliche Informationen für Services bei Unterforderung

Tabelle 10: konzeptuelle Ansätze zur Unterstützung der Lotsenarbeit und Vermeidung von Überlastsituationen hervorgehend aus den Voruntersuchungen

Das iterative Vorgehen umfasste dabei vier wesentliche Stufen: Zunächst wurden aus den Vorstudien heraus eine Vorstellung über den entsprechenden Aspekt entwickelt, sowie Anforderungen an die Umsetzung des Aspektes. Aus diesen Vorstellungen wurden in einem nächsten Schritt mehrere Design- und Konzeptideen entwickelt. Dabei wurde viel Wert auf die Unterschiedlichkeit der Entwürfe gelegt und als formales Hilfsmittel wurden oft gemeinsame Zeichen-Sessions zur Ideenfindung genutzt. Diese Ideen wurden anschließend prototypisch umgesetzt. In den ersten Iterationen oft als einfache Papierprototypen oder Zeichnungen, welche in jeder Iteration bis zu einem technischen Prototypen verfeinert wurden. Der Anschließende Evaluationsschritt, in welchem Techniken wie Fokusgruppen, Interviews und Online-Experimente zum Einsatz kamen, brachte wiederum neue Einsichten in die entsprechende Thematik und induzierte eine Überarbeitung unserer Vorstellungen und Anforderungen.

Das finale Designkonzept der adaptiven Nutzerschnittstelle

In diesem Abschnitt wird das Design des Fluglotsenarbeitsplatzes betrachtet. Zunächst werden jene Designaspekte betrachtet, die vorbeugend gegen kritische Situationen das Design des Arbeitsplatzes beeinflussen. Anschließend werden das Design der Stressvisualisierung und die adaptiven Aspekte des Designs näher erläutert.

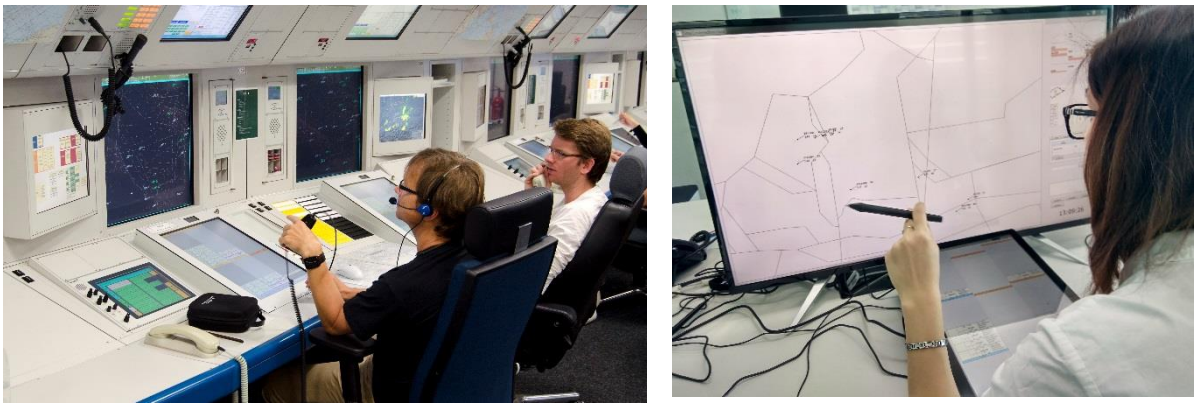


Abbildung 53: links: Das aktuelle Lotseninterface im Center in München (Quelle: DFS Deutsche Flugsicherung GmbH), rechts: unser Prototyp für ein Lotseninterface

Um einer stressigen Situation vorzubeugen sollen die doch sehr vielen Daten in eine übersichtlichere Darstellungsform gebracht werden. Dazu wurden Designentscheidungen zugunsten kleinerer Veränderungen im Gesamtdesign getroffen. Ein einheitliches Design kann mehr Struktur ausstrahlen. Aufgrund dessen entschlossen wir uns zunächst sämtliche Darstellungen in einer Positivdarstellung anzuzeigen. Dies sollte zudem für eine weniger starke Ermüdung des Auges führen, da der Wechsel zwischen hellen und dunklen Hintergründen anstrengender für das Auge ist [7]. Zudem sollten die vielen Bildschirme, die den Arbeitsplatz insgesamt sehr unruhig erscheinen lassen, auf möglichst wenige Bildschirme reduziert werden. So sollen in unserem Prototyp sämtliche Bildschirme für Sekundärinformationen mit dem Radardisplay zusammengefasst werden. Die flugzeugbezogenen Daten (Rufzeichen, Position, Heading, Steigraten, Geschwindigkeit usw.) bleiben am wichtigsten und sollen daher stets auf der Radaransicht verfügbar sein. Andere situativ relevante Informationen (z.B. Wetteranomalien) sollen in kleinen Übersichtsansichten in einer Seitenleiste zur Verfügung stehen, die bei Bedarf kurzzeitig in die Radaransicht integriert werden können. Dadurch können die Fluglotsen den räumlichen Bezug beispielsweise zwischen Wetterdaten und der Flugsituation einfacher herstellen.

Eine dieser Sekundäransichten wird die neue Stressvisualisierung, welche die Daten aus dem MACeLot-System bezieht. Eine durchgeführte Fokusgruppe mit 4 Fluglotsen im Mai 2016 enthüllte einige Anforderungen an diese Übersicht. Sie sollte auf einen Blick eine gute Übersicht über die Nachbarsektoren bieten und eine Referenzierung der Sektoren über ihr Kürzel ist notwendig. Von den vorgelegten Skizzen kamen insbesondere die Anordnung über eine Liste als auch ein nach geografischem Vorbild angeordnete Darstellung gut an. Diese wurden für den finalen Prototypen weiter verfeinert.

In kritischen Ausnahmesituationen soll sich die Anzeige folgendermaßen automatisch an die spezifische Situation anpassen: Die weiterhin unverzichtbaren wichtigen Informationen werden auf die gleiche Weise wie zuvor zugänglich sein. Insgesamt wird aber die visuelle Komplexität in Ausnahmesituationen reduziert. Wir konnten feststellen, dass in Hochlastsituationen insbesondere jene Informationen weniger benötigt werden, die dem „Service-Gedanken“ dienen. Dies sind beispielswiese Serviceinformationen an die Piloten, die Abkürzungen ermöglichen (kommende Sektorenfolge, Richtung des Zielflughafens usw.), aber auch Serviceinformationen im Centerbetrieb wie zum Beispiel der Stresslevel der Kollegen. Die Stressvisualisierung angeordnet in der Seitenliste kann jederzeit bei Bedarf kon-

sultiert werden. Zudem haben wir in unserem Prototypen die Idee verfolgt wichtige (Zwischen-)Zielpunkte eines Fliegers durch eine Pfeilspitze um das Flugzeugsymbol in Unterlastsituationen automatisch einzublenden.

Eine weitere Erkenntnis unserer Untersuchungen ist, dass die Flugzeuge bei steigendem Stress zunehmend als gleichartig wahrgenommen werden: die Wichtigkeit von Informationen wie der Flugzeugtyp, die Wirbelschleppenkategorie oder das Equipment nimmt ab. Solche Informationen können in Ausnahmesituationen ausgeblendet werden und nur noch auf Abruf zur Verfügung stehen. Da ein komplettes Ausblenden von Informationen in einem sicherheitskritischen Bereich in Einzelfällen durchaus zu einem unnötigen Risiko führen kann, entschlossen wir uns lediglich für eine optische Zurücknahme der Informationen auf den Flugstreifen durch Ausgrauen.

Eine weitere Idee zur Adaption bei Stress beruht auf der Annahme, dass sich bei einer anhaltenden Überlastung die Arbeitsweise der Lotsen verändert: anstatt vorausplanend zu arbeiten, verfahren sie zunehmend reaktiv. Eine Erhöhung des Automatisierungsgrades kann dazu dienen, die Lotsen während ihrer reaktiven Arbeitsweise zu unterstützen. Die Kollegen an der ENAC in Frankreich untersuchten diesbezüglich zum Beispiel die Anpassung von Alarmen [59] [60]. Eine entsprechende Verwendung dieser Forschungsergebnisse für unser System erscheint sinnvoll.

Evaluierung der adaptiven Nutzerschnittstelle anhand eines technischen Prototypen

Die vorerst finale Betrachtung des entstandenen Designkonzeptes fand im Dezember 2017 am DFS-Standort München statt. Hierbei wurden 15 Lotsen mit einem technischen Prototyp konfrontiert, der die zuvor beschriebenen Designaspekte (das allgemeine Arbeitsplatzdesign, die Darstellung des Belastungszustandes der umliegenden Sektoren, die Wetterdarstellung und die automatische Adaption im Falle von Über- und Unterlast) exemplarisch umsetzte. Die Lotsen durften diesen Prototypen spielerisch erkunden und dabei Fragen stellen. Sie wurden aufgefordert währenddessen ihre Gedanken, Hoffnungen und Ängste zu äußern. Anschließend wurde sie noch einmal speziell zu den noch nicht entdeckten Aspekten und deren Vor- und Nachteilen befragt. Die Daten wurden mittels Diktiergerät aufgenommen und zur Auswertung zunächst transkribiert, anschließend codiert und kategorisiert.

Es werden im Folgenden zunächst die Angaben über die Änderungen am allgemeinen Design und anschließend die Adaptionen im Falle von Über- oder Unterlast der Fluglotsen erläutert.

Die Darstellung aller Ansichten in der Positivdarstellung wurde von den Lotsen sehr positiv kommentiert. Sie halten diese Ansicht für weniger anstrengend und sie loben, dass man durch den hellen Hintergrund die Reflexionen von Licht auf der Bildschirmoberfläche weniger wahrnimmt und dadurch die Helligkeit im Center erhöht werden könnte. Die Zusammenfassung der Daten auf einen großen Bildschirm wurde sowohl mit Vor- als auch mit Nachteilen gesehen. Allgemein wirkt nach Aussage der Fluglotsen ein einzelner Bildschirm aufgeräumter, die Informationen fallen schneller ins Auge und es sind weniger Kopfbewegungen nötig, was sich zeitlich positiv auswirkt. Hingegen wurde befürchtet, dass man durch einen so großen Bildschirm zu weit weg vom Partner sitzen würde oder auf diesem Bildschirm stetig den Mauszeiger suchen würde. Ersterem Nachteil kann bereits im jetzigen Prototyp entgegengewirkt werden, indem die Seitenleiste mit den Sekundärinfos wahlweise an den rechten oder linken Bildschirmrand verlegt werden kann. Der andere Nachteil könnte durch die zusätzliche Touch- oder Stifteingabe auf dem Radarbildschirm gelöst werden. Die Sidebar wurde in dem Zusammenhang positiv gesehen, da sie die Übersichtlichkeit des Interfaces begünstigt. Es wurde der Wunsch geäußert, die Sidebar modularisiert und in der Größe flexibel zu halten. Das heißt, dass die Seitenleiste

nach Wunsch mit den passenden Informationsdarstellungen gefüllt werden kann und sowohl die Größe der Module als auch der gesamten Leiste veränderbar ist.

Unsere Seitenleiste beinhaltete im Prototyp eine vereinfachte Wetterdarstellung (Abbildung 54), mit der Möglichkeit für ein Overlay auf das Radarbild. Die vereinfachte Winddarstellung wurde im Gegensatz zu heutigen Ansichten positiv wahrgenommen. Das Overlay kam in einigen Fällen sehr gut an (SFluglotse 14: „Das wäre perfekt!“), wurde aber teilweise auch als unnötig empfunden. Da dies ein optional anzeigbares Feature ist, sollte es demnach beibehalten werden.

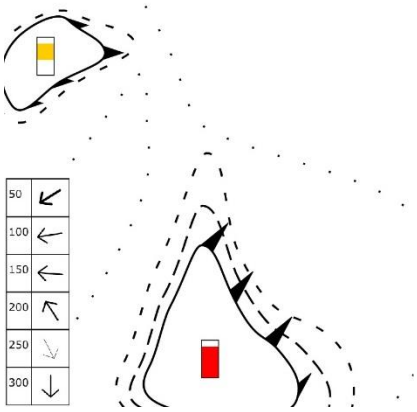


Abbildung 54: Vereinfachte Wetterdarstellung: Windrichtung und -geschwindigkeit werden in einer Tabelle für den gesamten Sektor angegeben

Die automatische Stressanzeige wurde von allen Fluglotsen begrüßt. Diese Anzeige würde das Problem lösen, dass Fluglotsen im Fall von hohen Stress vergessen könnten, ihren Status anzugeben. Die Anordnung der Sektoren kam hälftig mit räumlicher Anordnung, hälftig als alphabetisch und nach Höhe sortierte Liste gut an. Eine geografische Anordnung, die schnell die Richtung des Stresses angibt, wurde als Vorteil für Neulinge empfunden. Es bestand jedoch die Befürchtung, dass, wenn man eine Zulassung auf mehreren Sektoren hat, stets die Sektoren suchen würde. Da die Darstellung insgesamt möglichst kompakt sein sollte, wurde von den Lotsen eine farbige Fläche pro Sektor, die ihren Farbwert anpasst als vorteilhaft empfunden. Als Verbesserungen wurde die Kopplung mit der Frequenzübersicht, dem Abgleich der Anordnung mit den Telefontasten und die automatische Vorhersage von und Warnung vor Stresssituationen genannt.

In Bezug auf die automatische Adaption ihrer Interfaces sind die Fluglotsen eher zurückhaltend bis strikt ablehnend. Sie befürchten, dass die Veränderung in der Darstellung ihre Routinen durchbricht und sie somit stark in ihren Kapazitäten gebunden wären. So wurde auch die Idee der Einblendung von Servicedaten bei Unterforderung durchweg abgelehnt. Hier kam zu der Angst vor der Veränderung zusätzlich zum Tragen, dass man sich unter Druck gesetzt fühlen könnte alle Vorschläge zu erfüllen. Dies würde wiederum Stress erzeugen sowie der Empfindung der Fluglotsen entgegenstehen, dass diese Einblendung unnötig sei, da ein erfahrener Lotse die groben Richtungen der Zielflughäfen oder mögliche Abkürzungen bereits aus der Erfahrung herauskennt. Als Alternative zur Aufrechterhaltung der Konzentration wurde vorgeschlagen, die Ankündigung der Flugzeuge durch ihre Flugstreifen in Abhängigkeit von der Unterforderung bereits früher einzublenden oder in regelmäßigen Abständen eine Nutzerinteraktion zu verlangen.

Aufgrund der grundsätzlichen Abneigung bezüglich der Adaptionen war es umso mehr erstaunlich, dass die Idee der Zurücknahme von Informationen auf den Streifen teilweise positiv aufgefasst wurde.

Hervorgehoben wurde insbesondere der verbesserte Überblick über die vorhandenen Streifen in besonderen Situationen: Solche in denen man einerseits aufgrund der Verkehrsmenge viel zu viele Streifen auf dem Streifenboard und andererseits auch viel länger den Blick auf dem Radarbild hat. Gleichzeitig wurde klar, dass die Auswahl der auszugrauenden Informationen feinfühliger getroffen werden muss. Entweder jeder Lotse kann diese individuell festlegen oder es müssen umfangreichere Studien als unsere initiale Umfrage durchgeführt werden. Hier würde sich eine Kombination aus Umfragen mit einer weitaus größeren Stichprobe als der unsrigen und objektiven Beobachtungen und Messungen anbieten. Zudem wäre eine Auswertung der Veränderungen basierend auf der Messung der Performanz der Lotsen anstelle der individuellen Eindrücke sinnvoll, um eine Effektivität der Adaption festzustellen. Auch im Falle der Überlast wurden Alternativen und Erweiterungen des Konzeptes vorgeschlagen. So wünschten sich die Lotsen beispielsweise lieber eine Hervorhebung der Callsigns anstelle des Ausgrauens der übrigen Informationen; in Hochlastsituationen einen Vermerk der Frequenz des Nachfolgesektors auf dem Streifen, sodass sie diese bei der Verabschiedung des Flugzeugs nicht suchen müssen; eine Hervorhebung der Flieger zu denen der Lotse sich Freigaben geholt hatte oder das Ausblenden von überflüssigen Targets.

Insgesamt wurde von den Fluglotsen viel Wert daraufgelegt, dass ein stressreduzierendes Interface in seiner Gestaltung Stress von vorneherein reduziert. Demnach wurden die vorgeschlagenen Änderungen im Gesamtdesign größtenteils positiv aufgenommen. Die automatischen Adaptionen hingegen eher abgelehnt. Dennoch erwies sich die Idee der Zurücknahme von Informationen in Stresssituationen als Idee, die weiter untersucht werden sollte.

Visualisierungsanalyse des Anwendungsfeldes und des Emotionsmodells für die Wachleiter

Als Nutzer einer Visualisierung des Emotionsmodells stehen die Wachleiter (Supervisoren) der Fluglotsen im Mittelpunkt. Eine Verwendung solcher personenbezogener Daten (z.B. von Managern) außerhalb des Betriebsraumes wurde von den Lotsen und Supervisoren aus persönlichkeitsrechtlichen Gründen strikt abgelehnt und deswegen auch nicht weiter verfolgt.

Ebenso wie für die Nutzerschnittstelle der Fluglotsen starteten wir auch für die Entwicklung dieser Visualisierung mit der Analyse der Arbeitsweise der Wachleiter. Eine zweitägige Beobachtung der Tätigkeiten der Supervisoren im DFS Center München (September 2016) gab Einblicke in die Ziele und entwickelten Arbeitsweisen. Interviews mit insgesamt sieben Wachleitern sollten zudem einen tieferen Einblick in die Entscheidungsprozesse der Wachleiter und deren Informationsbedarf bezüglich des Stresses und der emotionalen Verfassung der Lotsen geben.

Die Daten der Beobachtungen wurden mittels Stift und Papier aufgenommen, während von den Interviews Audioaufzeichnungen gemacht wurden, welche transkribiert für die Auswertung zur Verfügung standen. Die Daten wurden in einem Bottom-Up-Prozess kodiert und kategorisiert. Anschließend wurden aus den gewonnenen Erkenntnissen Designanforderungen für die Visualisierung abgeleitet.

Ziel der Supervisoren ist es, den „Betrieb am Laufen zu halten“ (Wachleiter 2). Dies bedeutet, alles zu tun, damit die Umstände es den Fluglotsen erlauben den Flugverkehr sicher und flüssig durch den Luftraum zu geleiten. Diese Aufgabe beinhaltet hauptsächlich die Erstellung eines Dienstplanes, welcher die Lotsen den Arbeitspositionen zuweist, sowie das Lösen jeglicher Probleme und Engpässe, die die Sicherheit des Flugverkehrs gefährden könnten. Tagtäglich wird der Wachleiter somit mit komplexen Problemen konfrontiert [43], [119]. Er adressiert viele unterschiedliche Variablen, die alle miteinander verknüpft sind, seine Zeit zum Treffen von Entscheidungen ist oftmals zeitlich stark begrenzt und es

tauchen regelmäßig unerwartete Ereignisse (z.B. Notfälle, externe Anforderungen oder Häufungen von Krankheitsfällen) auf. Ein Wachleiter muss unterschiedliche Ziele (Sicherheit, Kosteneffizienz und die Zufriedenheit der Fluglotsen) bei seinen Entscheidungen berücksichtigen und die Informationen, die er dazu zur Verfügung hat sind oft unvollständig und nur bedingt zuverlässig. Er trifft Entscheidungen für die Zukunft auf Basis von aktuellen Daten, einer aus der Erfahrung gewachsenen persönlichen Heuristik und unzuverlässigen Vorhersagen (z.B. Wetter oder Verkehrsquantität). Hinzu kommt, dass es viele unterschiedliche Lösungswege aber keine definierte optimale Lösung gibt. Dieser Problemtyp lässt sich aufgrund der Unvorhersehbarkeit und Komplexität der Situationen schlecht automatisieren, aber sehr gut durch den Computer (zuverlässige Vorhersagealgorithmen und gute visuelle Darstellungen) unterstützen.

Auch in der heutigen Arbeit ist eine der Variablen, die für die Entscheidungen der Supervisoren herangezogen wird, die Verfassung der Fluglotsen. Dies beinhaltet beispielsweise ihre tagesaktuelle Leistungsfähigkeit und Stimmung, ihren Erschöpfungszustand oder ihre Zufriedenheit. Diese Informationen werden in erster Linie über die interpersonelle Kommunikation und aufmerksame Beobachtung erlangt. Die Vorhersage von Stress wird zudem durch die Vorhersage der Verkehrsquantität im Sektor unterstützt. Der Zugang zu diesen Daten kann durch die automatische Messung und Vorhersage von kognitiver Arbeitslast, Arousal (Erregtheit) und emotionaler Valenz sinnvoll erleichtert werden.

Des Weiteren konnten aus den Beobachtungen und Interviews Anforderungen an die Visualisierung dieser Werte festgestellt werden. Die von den Supervisoren genutzten Planungsintervalle und ihre benötigten Reaktionszeitenführen direkt zu einer dieser Anforderungen: :

R1: Die Werte zum Lotsenzustand sollten in einem Zeitraum von zwei Stunden in der Vergangenheit bis zwei Stunden in der Zukunft verfügbar sein.

Die wichtigste Information für den Wachleiter ist dabei, ob der Fluglotse arbeitsfähig oder arbeitsunfähig ist. Die Verwendung von automatisch erkannten Emotionsdaten wird allerdings von den Supervisoren strikt abgelehnt. Als Gründe wurden benannt: der Schutz der jeweiligen Person und ihrer Privatsphäre sowie die Angst unangemessene Entscheidungen durch Einbeziehung der Emotionsdaten zu treffen. Dies führt zu zwei weiteren Anforderungen an die Visualisierung:

R2: Emotionsdaten sollten weitestgehend unzugänglich sein. Lediglich der Arousalwert, der zwar einen Hinweis auf den Emotionszustand geben kann, aber auch einen starken Hinweis über die Arbeitsfähigkeit des Lotsen gibt, soll Verwendung finden.
R3: Extreme Situationen sollen sofort ins Auge fallen. Extremer Stress, Langeweile oder eine starke Erregung können den Fluglotsen in seiner Arbeit behindern und eine Intervention des Wachleiters kann angebracht sein.

Zwei der zentralen Aufgaben des Wachleiters sind zum einen die Suche nach einem Lotsen, der eine bestimmte Arbeitsposition übernehmen kann, und zum anderen die Suche nach möglichen Engpässen in den Sektoren. Entsprechend sollten die Daten (kognitive Belastung und Arousal) im Bezug zu beiden strukturellen Kategorien vorliegen. Daraus ergibt sich folgende vierte Anforderung:

R4: Die mentalen Modelle der Wachleiter sollten berücksichtigt werden. Die Daten sollen in Bezug auf den Sektor und in Bezug auf den Lotsen zugänglich sein.

Neben den aus der Vorstudie entwickelten Anforderungen sollte die Nutzerschnittstelle ebenfalls die Anforderungen erfüllen, die in Bezug auf die menschlichen Fähigkeiten in der Kognition und Wahrnehmung sinnvoll erscheinen.

R5: Es soll lediglich eine minimale Anzahl an Primitiven verwendet werden, um eine aussagekräftige und effektive Visualisierung zu schaffen [143]. Alle wichtigen Informationen sollen einfach identifizierbar sein und alle visuellen Elemente sollen eine Bedeutung haben.

R6: Der Einsatz von Farbe soll auf die wichtigsten Elemente beschränkt werden und die visuellen Wahrnehmungsfähigkeiten des Menschen berücksichtigen [162].

Die Entwicklung einer einfachen und übersichtlichen Darstellung scheint insbesondere unter dem Aspekt, dass zahlreiche Informationen stetig durch die Wachleiter abgerufen werden müssen sinnvoll. Schon heute besteht ihr Arbeitsplatz aus einer Vielzahl an Bildschirmen und Anzeigen, demnach sollte sich die Visualisierung am besten nahtlos in die bestehenden Tools einfügen. Eine ausführliche Abhandlung über die Vorstudie ist in [111] zu finden.

Interaktions- und Visualisierungsdesign

Der Designprozess verlief iterativ unter stetiger Einbeziehung der zukünftigen Nutzer. Es wurden Designideen generiert und Prototypen entworfen. Die Prototypen wurden evaluiert, woraufhin die Designanforderungen angepasst und die Designideen verworfen oder abgeändert wurden. Im Verlauf des Prozesses nahm die Granularität der Prototypen zu. Den Wachleitern wurden anfänglich einfache Skizzen und gegen Ende des Projektes ein technischer Prototyp präsentiert.

Die erste Designentscheidung betraf die Aufteilung der Darstellung in eine Übersichtsvisualisierung und eine Detailansicht. Die Übersicht (Abbildung 55) ist dabei stets sichtbar und bietet einen schnellen Überblick über die Gesamtsituation im Center, wobei sie klar die kritischen Situationen kommunizieren soll (R3). Die Detailansicht (Abbildung 57) soll lediglich bei Bedarf abrufbar sein und dient der genaueren Analyse einer kritischen Situation.

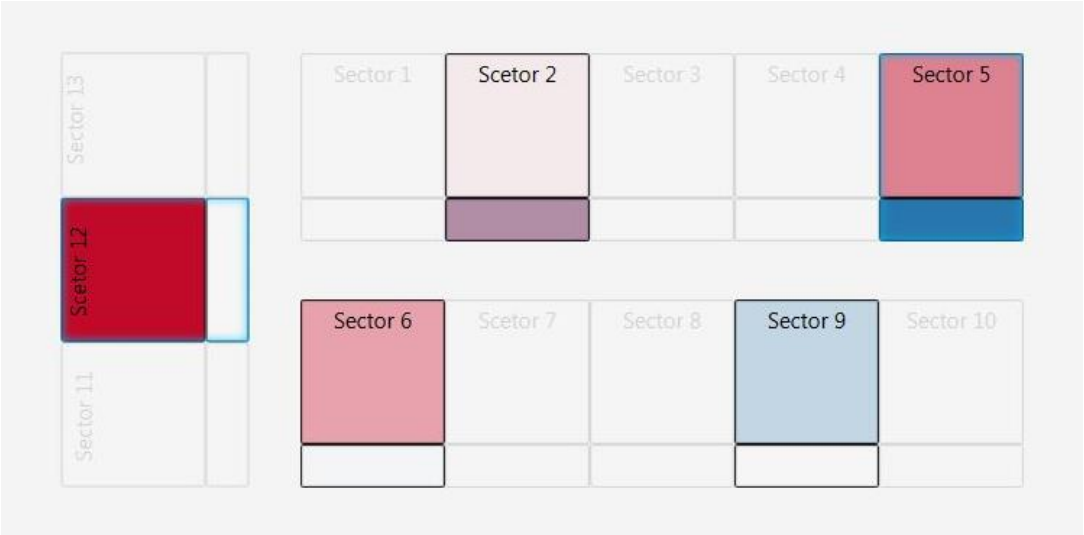


Abbildung 55: die Übersichtsvisualisierung dient dem schnellen Überblick der Situation im Center.

Die Anordnung der dargestellten Werte entspricht der Anordnung der Arbeitsplätze im Center. Dies soll das Mapping zwischen abstrakten Bezeichnungen und den mentalen Modellen der Supervisoren unterstützen. Die Verwendung von Farbmarkierungen wurde auf die wichtigste Information

beschränkt, nämlich der Existenz von kritischen Situationen (R6), somit ist auch eine periphäre Wahrnehmung kritischer Situationen möglich (R3). Die Farbskala reicht von Blau (Unterlast) über ein helles Grau (Komfortzone) hin zu einem Rot (Überlast). Die Darstellung sollte möglichst einfach sein (R5), dabei aber gleichzeitig nicht nur die aktuelle sondern auch die Vorhersage der Stress- und Arousalwerte berücksichtigen. Die angezeigte Farbe ist demnach ein vom aktuellen Zeitpunkt bis in die Zukunft aufsummierter Wert, wobei der aktuelle Zeitpunkt mit 100 % Opazität beiträgt und die zukünftigen Werte mit zunehmender Durchsichtigkeit aufaddiert werden (Abbildung 57). Beim Hovern mit der Maus über das jeweilige Sektorsymbol werden zusätzlich im sich öffnenden Tooltip auch die individuellen Werte der dort aktuell arbeitenden Lotsen angezeigt (R4). Eine kleinere Version der farbigen Flächen kann der im Planungstool bereits existierenden Lotsenliste hinzugefügt werden (Abbildung 56), womit auch der stetige Zugang zu den individuellen Werten (R4), sowie die mögliche Integration in die bestehenden Tools und Arbeitsabläufe sicher gestellt ist.

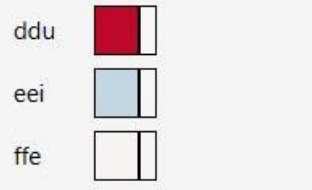


Abbildung 56: Übersicht der Lotsenbezogenen Stress- und Arousalwerte in einer Liste mit Lotsenkürzeln.

Die Detailansicht (Abbildung 57), die bei Bedarf für ausgewählte Sektoren oder Lotsen aufgerufen werden kann, berücksichtigt nun den gesamten zeitlichen Verlauf der Werte in der Vergangenheit und Zukunft (R1) in einem einfachen Liniendiagramm. Diese sind durch ihre weite Verbreitung bei der Darstellung zeitlicher Verläufe von vielen Nutzern gut interpretierbar. Auch hier wird die oben beschriebene Farbskala zur Hervorhebung der Extrema verwendet. Zudem kann die Zusammensetzung der aufsummierten Farbwerte aus der Übersicht (hier in der Mitte abgebildet) nachvollzogen werden.

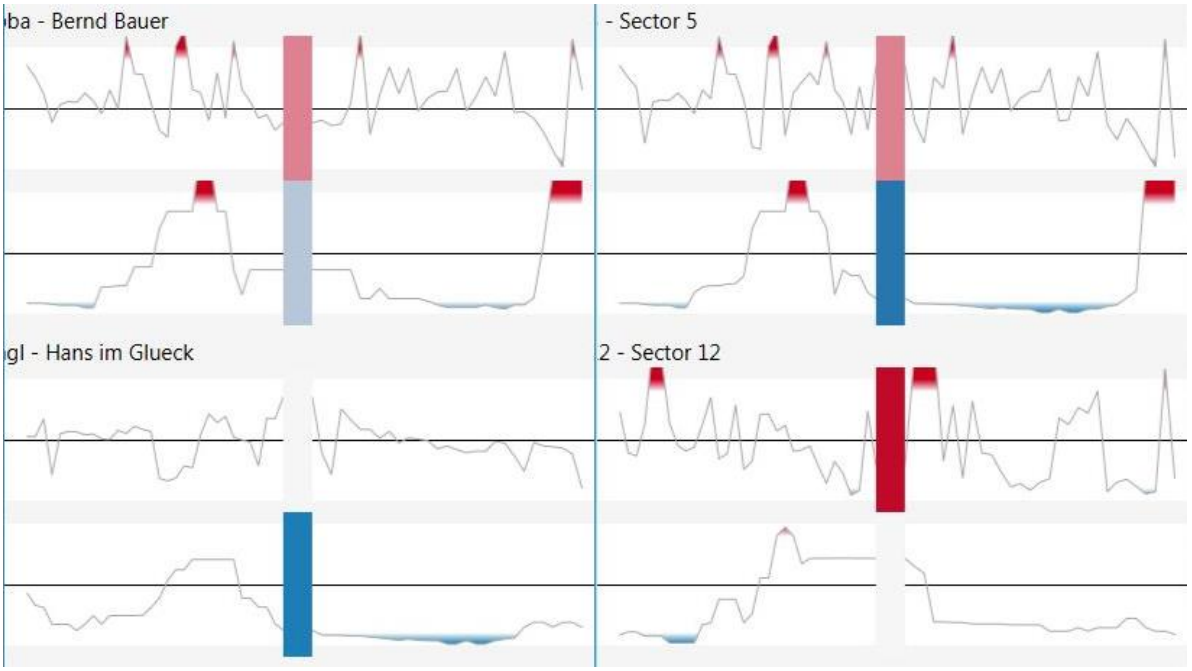


Abbildung 57: Die Detailansicht zeigt ausgewählte Arbeitslast- und Arousalwerte in ihrem zeitlichen Verlauf.

2. Voraussichtlicher Nutzen, insbesondere der Verwertbarkeit des Ergebnisses im Sinne des fortgeschriebenen Verwertungsplans

Die größte Anschlussfähigkeit in wirtschaftlicher Hinsicht besteht naturgemäß in Bezug auf den Anwendungsfall des Lotsenarbeitsplatzes. Im Zusammenhang mit dem Anwendungspartner wurden verschiedene Ansätze zur emotions- und kommunikationssensitiven Anpassung der Nutzerschnittstelle, insbesondere der Visualisierung, entwickelt. Die Untersuchung der technischen Umsetzbarkeit dieser Ansätze, der Nutzerakzeptanz und der Auswirkung auf die Arbeit der Lotsen war Bestandteil dieses Projekts.

Der ELSI-Workshop im Jahr 2016 diente zur Exploration der Nutzerakzeptanz und zur Diskussion der möglichen Auswirkungen des MACeLot-Ansatzes auf die Arbeit der Fluglotsen. Die am Workshop teilnehmenden Fluglotsen sahen das größte Potenzial für den Einsatz der Software in den von der DFS regelmäßig durchgeführten Simulationsübungen, da unsere Software ein objektives Messwerkzeug darstellt und zuverlässigere Ergebnisse erwarten lässt als die bisherige Methode der händischen Bewertung des emotionalen Zustands eines Mitarbeiters.

Es bestehen sehr gute Anschlussmöglichkeiten im Hinblick auf die Entwicklung eines produktiv einsetzbaren Assistenzsystems auf der Grundlage der gewonnenen Erkenntnisse. Die vorgesehene Visualisierung der Zustandsdaten lässt sich unmittelbar auf andere Anwendungen übertragen, bei denen Supervisoren zur Unterstützung von Operateuren eingesetzt werden. Die wissenschaftlichen Erkenntnisse und entwickelten Methoden der Emotionserkennung haben grundlegende Bedeutung für die Entwicklung von emotionssensitiven Systemen im Allgemeinen und in Bezug auf Teamarbeitsplätze im Besonderen.

Die Projektergebnisse wurden dem Fachpublikum bereits auf Konferenzen und Workshops präsentiert und entsprechend publiziert.

3. Während der Durchführung des Vorhabens dem ZE bekannt gewordenen Fortschritts auf dem Gebiet des Vorhabens bei anderen Stellen

Die Entwicklung emotionssensitiver Systeme ist ein hochaktueller Forschungsbereich, der gegenwärtig in verschiedenen Forschungsprojekten bearbeitet wird. Neben den Projekten des BMBF-Programms „InterEmotio“ sind das EU Projekt **„Human Performance Neurometrics toolbox for highly automated systems“** (STRESS, http://www.stressproject.eu) oder die **ESF-Nachwuchsforschergruppe „Sozial agierende, kognitive Systeme zur Feststellung von Hilfsbedürftigkeit”** (https://www.tu-chemnitz.de/informatik/KI/projects/social/) Beispiele hierfür. Dementsprechend werden laufend neue Forschungsergebnisse in diesem Bereich publiziert. Eine Methodenbasis zur Entwicklung von Assistenzsystemen für Fluglotsen, die in Konkurrenz zu dem in diesem Projekt erzielten Ergebnisse steht, ist jedoch nicht publiziert worden.

4. Erfolgte oder geplante Veröffentlichungen der Ergebnisse

Im Projektzeitraum wurden folgende Publikationen veröffentlicht:

[Pub1] Buxbaum, J., Müller, N. H., Ohler, P., Pfeiffer, L., Rosenthal, P., Valtin, G. (2016). Emotion-sensitive automation of air traffic control - Adapting air traffic control automation to user emotions. In International Transportation, 68 (1), pp. 36-39. ISSN: 0020-9511

[Pub2] Ebersbach, M., Herms, R., Eibl, M. (2017). Fusion Methods for ICD10 Code Classification of Death Certificates in Multilingual Corpora. In CLEF 2017 Evaluation Labs and Workshop: Online Working Notes, Dublin, Ireland. CEUR-WS. ISSN: 1613-0073

[Pub3] Ebersbach, M., Herms, R., Lohr, C., Eibl, M. (2016). Wrappers for Feature Subset Selection in CRF-based Clinical Information Extraction. In CLEF 2016 Evaluation Labs and Workshop: Online Working Notes, pp. 69--80, Évora, Portugal. CEUR-WS, ISSN: 1613-0073

[Pub4] Herms, R. (2016). Prediction of Deception and Sincerity from Speech Using Automatic Phone Recognition-Based Features. In Proceedings of INTERSPEECH 2016, pp. 2036-2040, San Francisco, CA, USA. ISCA. ISSN: 1990-9772

[Pub5] Herms, R., Seelig, L., Münch, S., Eibl, M. (2016). A Corpus of Read and Spontaneous Upper Saxon German Speech for ASR Evaluation. In Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016), Portorož, Slovenia. ELRA. ISBN: 978-2-9517408-9-1

[Pub6] Herms, R., Richter, D., Eibl, M., Ritter, M. (2015). Unsupervised Language Model Adaptation using Utterance-based Web Search for Clinical Speech Recognition. In CLEF 2015 Evaluation Labs and Workshop: Online Working Notes, Toulouse, France. CEUR-WS. ISSN: 1613-0073

[Pub7] Kowerko, D., Rößner, M., Kahl, S., Herms, R., Eibl, M., Engelmann, K. (2017). Aufbereitung augenmedizinischer Bild-, Patienten- und Diagnosedaten zum Zwecke der Forschung - Ethikrichtlinen und deren praktische Umsetzung. In Mensch und Computer 2017 - Workshopband, Regensburg, Germany. GI. DOI: 10.18420/muc2017-ws07-0288

[Pub8] Lohr, C., Herms, R. (2016). A Corpus of German Clinical Reports for ICD and OPS-based Language Modeling. In Proceedings of the Controlled Language Applications Workshop (CLAW) at LREC 2016}, pp. 20--23, Portorož, Slovenia. ELRA. ISBN: 978-2-9517408-9-1

[Pub9] Müller, N. H., Truschzinski, M. (2015). Analytical Steps for the calibration of an Emotional Framework – Pre-Test and Evaluation Procedures. In: Human-Computer Interaction: Design and Evaluation Lecture Notes in Computer Science Volume 9169, 2015, pp 512-519. ISBN: 978-3-319-20901-2

[Pub10] Pfeiffer, L, Valtin, G., Müller, N.H., Rosenthal, P. (2016): The Mental Organization of Air Traffic and its Implications to an Emotion Sensitive Assistance System. In International Journal on Advances in Life Sciences, 8 (1&2), S. 164-174. ISSN: 1942-2660

[Pub11] Pfeiffer, L., Müller, N. H., Valtin, G., Truschzinski, M., Ohler, P., Rosenthal, P. (2015). Emotionsmodell für zukünftige Mensch-Technik-Schnittstellen zur Unterstützung von Centerlotsen. DGLR-Bericht 2015-01. ISBN: 9783932182839 bzw. ISSN: 0178-6326

[Pub12] Pfeiffer, L., Müller, N. H., Rosenthal, P. (2015). A Survey of Visual and Interactive Methods for Air Traffic Control Data. Proceedings of the International Conference on Information Visualisation, pp. 574–577. DOI: 10.1109/iV.2015.102

[Pub13] Pfeiffer, L., Sims, T., Rosenthal, P. (2017). Visualizing Workload and Emotion Data in Air Traffic Control – An Approach Informed by the Supervisors Decision Making Process. In ACHI 2017, The Tenth International Conference on Advances in Computer-Human Interactions, S. 81-87. ISBN: 978-1-61208-538-8

[Pub14] Pfeiffer, L., Valtin, G., Müller, N. H., Rosenthal, P. (2015). Aircraft in Your Head: How Air Traffic Controller Mentally Organize Air Traffic. Pascal Lorenz, Christian Bourret (Ed.): Proceedings of HUSO, the International Conference on Human and Social Analytics, pp. 19–24, IARIA (Best Paper Award). ISBN: 9781612084473

[Pub15] Pfeiffer, L., Valtin, G., Müller, N.H., Rosenthal, P. (2016). The Mental Organization of Air Traffic and its Implications to an Emotion Sensitive Assistance System. In: International Journal on Advances in Life Sciences, 8 (1&2), pp. 164-174.

[Pub16] Rosenthal, P., Pfeiffer, L., Müller, N.H., Valtin, G. (2016). The Long Way to Intuitive Visual Analysis of Air Traffic Control Data. In Elliot Bendoly, Sacha Clark (Ed.): Visual Analytics for Management: Translational Science and Applications in Practice, Chapter 11, Routledge. ISBN: 9781317278375

[Pub17] Truschzinski, M. (2017). Modeling Workload: A System Theory Approach. In Proceedings of the Companion of the 2017 ACM/IEEE International Conference on Human-Robot Interaction, ACM Press, S.305-306.

[Pub18] Truschzinski, M., Pfeiffer, L., Valtin, G., (2017). ELSI-Aspekte in Forschungsverbünden. In: Burghardt, M., Wimmer, R., Wolff, C. & Womser-Hacker, C. (Hrsg.), Mensch und Computer 2017 - Workshopband. Regensburg: Gesellschaft für Informatik e.V.. DOI: 10.18420/muc2017-ws07-0378

[Pub19] Truschzinski, M., Wirzberger, M. (2017) A dynamic process model for predicting workload in an air traffic controller task. In Computational foundations of cognition: 39th Annual Meeting of the Cognitive Science Society (CogSci 2017), volume 1, pages 1224–1229, London, 2017. Curran Associates, Inc. ISBN: 978-0-9911967-6-0

[Pub20] Truschzinski, M., Klein, M. (2017). Modeling the enactive emotion theory: Methodological considerations. AISB 2017 Convention, Bath, UK, (S. 209-213).

[Pub21] Truschzinski, M., Betella, A., Brunnett, G., F. M. J. Verschure, P. (2018) Emotional and cognitive influences in air traffic controller tasks: An investigation using a virtual environment? Applied Ergonomics, Volume 69, May 2018, Pages 1-9. DOI: 10.1016/j.apergo.2017.12.019

[Pub22] Truschzinski, M., Valtin, G., Ohler, P. (submitted) Modeling mood changes within a cognitive demanding air traffic controller task. IEEE Transactions on Automatic Control, submitted.

[Pub23] Truschzinski, M., Valtin, G., H. Müller, N. (2017) Investigating the Influence of Emotion in Air Traffic Controller Tasks: Pretest Evaluation. In Don Harris, editor, Engineering Psychology and Cognitive Ergonomics: Performance, Emotion and Situation Awareness: 14th International Conference, EPCE 2017, Held as Part of HCI International 2017, Vancouver, BC, Canada, July 9-14, 2017, Proceedings, Part I, (pages 220–231). Springer International Publishing, Cham, 2017. ISBN 978-3-319-58472-0

[Pub24] Truschzinski, M., Klein, M. (2017). Modellierung und Vorhersage von mentaler Arbeitsbeanspruchung in einem Fluglotsenaufgabenexperiment. In: Eibl, M. & Gaedke, M. (Hrsg.), INFORMATIK 2017. Gesellschaft für Informatik, Bonn. (S. 2295-2300). DOI: 10.18420/in2017_230

[Pub25] Wirzberger, M., Truschzinski, M., Schmidt, R., Barlag, M. (2017). Computer Science meets Cognition. In: Eibl, M. & Gaedke, M. (Hrsg.), INFORMATIK 2017. Gesellschaft für Informatik, Bonn. (S. 2273-2277). DOI: 10.18420/in2017_227

Literaturverzeichnis

[1] Abadi, Martín, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S. Corrado, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Ian Goodfellow, Andrew Harp, Geoffrey Irving, Michael Isard, Rafal Jozefowicz, Yangqing Jia, Lukasz Kaiser, Manjunath Kudlur, Josh Levenberg, Dan Mané, Mike Schuster, Rajat Monga, Sherry Moore, Derek Murray, Chris Olah, Jonathon Shlens, Benoit Steiner, Ilya Sutskever, Kunal Talwar, Paul Tucker, Vincent Vanhoucke, Vijay Vasudevan, Fernanda Viégas, Oriol Vinyals, Pete Warden, Martin Wattenberg, Martin Wicke, Yuan Yu, and Xiaoqiang Zheng. TensorFlow: Large-scale machine learning on heterogeneous systems, 2015. Software available from tensorflow.org.

[2] Ahlberg J., CANDIDE-3 an updated parameterized face, Technischer Report, Dept. of Electrical Engineering, Linkoping University, Sweden, 2001.

[3] Aigner, W., Miksch, S., Schumann, H., Tominski, C., „Visualization of Time-Oriented Data“, Springer, 2011

[4] Amogh Gudi, H. E. (2015). Deep learning based facs action unit occurrence and intensity estimation. IEEE.

[5] Atrey, P.K., Hossain, M.A.,El Saddik, A., Kankanhalli, M.S. „Multimodal fusion for multimedia analysis: a survey“, Multimedia Systems, Bd. 16, Nr. 6, S. 345–379, Nov. 2010.

[6] Azad, P., Asfour, T., Dillmann, T. (2008): Robust Real-time Stereo-based Markerless Human Motion Capture. IEEE/RAS International Conference on Humanoid Robots (Humanoids). Daejeon, Korea, pp. 700-707.

[7] Batliner, Anton; Schuller, Björn; Seppi, Dino; Steidl,Stefan; Devillers, Laurence; Vidrascu,Laurence; Vogt,Thurid; Aharonson, Vered; Amir, Noam „The automatic recognition of emotions in speech.“ In: Emo-tion-Oriented Systems. Springer Berlin Heidelberg, 2011. S. 71-99.

[8] Bagassi, S., F. De Crescenzio, and F. Persiani, “Design and evaluation of a four-dimensional interface for air traffic control,” Proceedings of the Institution of Mechanical Engineers, Part G: Journal of Aerospace Engineering, vol. 224, no. 8, pp. 937–947, 2010.

[9] Belentschikow, V. & Müller, N. H. (2014). Online Observations and Virtual Ethnography – Ethical Issues of a Qualitative Experiment regarding Facebook: General Online Research, Köln 2014

[10] Berger, Arne; Knauf, Robert; Eibl, Maximilian; Marcus, Aaron (2011). Moody mobile TV: adding emotion to personalized playlists. In: ACM MobileHCI '11, Proc. of the 13th International Conference on Human Computer Interaction with Mobile Devices and Services, S. 623-628.

[11] Berggren, N., Koster, E. H. W. & Derakshan, N. (2012). The effect of cognitive load in emotional attention and trait anxiety: An eye movement study. Journal of Cognitive Psychology, 24(1), 79-91.

[12] Bin, Y. and Lugger, M.: "Emotion recognition from speech signals using new harmony features." Signal Processing 90.5 (2010): 1415-1423.

[13] Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. Journal of machine Learning research, 3(Jan), 993-1022.

[14] Boersma, Paul (2001). Praat, a system for doing phonetics by computer. *Glot International* **5:9/10**, 341-345.

[15] Bradley, M. M., & Lang, P. J. (1999). *Affective norms for English words (ANEW): Instruction manual and affective ratings* (pp. 1-45). Technical report C-1, the center for research in psychophysiology, University of Florida.

[16] Branch, R. M. (2010). Instructional Design: The ADDIE Approach. Springer: New York

[17] Brookings, J. B., Wilson, G. F. & Swain, C. R. (1996). Psychophysiological responses to changes in workload during simulated air traffic control. Biological Psychology. 42. Pp 361-377.

[18] Burkert, P., Trier, F., Afzal, M. Z., Dengel, A., & Liwicki, M. (2015). DeXpression: Deep Convolutional Neural Network for Expression Recognition.arXiv preprint arXiv:1509.05371

[19] Burkhardt, F.; Paeschke, A.; Rolfes, M.; Sendlmeier, W.; Weiss, B. „A database of German emotional speech.“ In: Proc. Of Interspeech. 2005. S. 1517-1520.

[20] Camp, G., Paas, F. Rikers, R. & van Merrienboer, J. (2001). Dynamic problem selection in air traffic control training: a comparison between performance, mental effort and mental efficiency. Computers in Human Behavior. 17(5-6). Pp 575-595.

[21] Carreira, J., Agrawal, P., Fragkiadaki, K., & Malik, J. (2015). Human Pose Estimation with Iterative Error Feedback. *arXiv:1507.06550 [cs]*. Abgerufen von http://arxiv.org/abs/1507.06550

[22] Castellano, G., Gunes, H., Peters, C., & Schuller, B. (2014). Multimodal affect recognition for naturalistic human-computer and human-robot interactions. *Invited Chapter for Handbook of Affective Computing*, 246-257.

[23] Castellano, G. Kessous, L., Caridakis, G. „Emotion Recognition through Multiple Modalities: Face, Body Gesture, Speech“, in *Affect and Emotion in Human-Computer Interaction*, Bd. 4868, C. Peter und R. Beale, Hrsg. Berlin, Heidelberg: Springer Berlin Heidelberg, 2008, S. 92–103.

[24] Chen, C.H., Wang, P.S.-P. Hrsg., *Handbook of pattern recognition and computer vision*, 3rd ed. River Edge, NJ: World Scientific, 2005.

[25] Cheron, G., Laptev, I., Schmid, C., 2015. P-CNN: Pose-Based CNN Features for Action Recognition. Presented at the Proceedings of the IEEE International Conference on Computer Vision, pp. 3218–3226.

[26] Christodoulidis, A., Delibasis, K.K., Maglogiannis, I. (2012): Near real-time human silhouette and movement detection in indoor environments using fixed cameras. PETRA '12, ACM 1-7

[27] Collet, C., Averty, P., Dittmar, A. Autonomic nervous system and subjective ratings of strain in air-traffic control, Applied Ergonomics, Volume 40, Issue 1, January 2009, Pages 23-32, ISSN 0003-6870

[28] Cootes T.F., Taylor C.J., Statistical Models of Appearance for Computer Vision, Technischer Report, University of Manchester, 2004.

[29] Cui Y., Jin Z., Facial Feature Points Tracking Based on AAM with Optical Flow Constrained Initialization, Journal of Pattern Recognition Research, vol.78(1): 72-79, 2012.

[30] Dang, N.T., Air Traffic Control. InTech, 2010, ch. Investigating Requirements for the Design of a 3D Weather Visualization Environment for Air Traffic Controllers

[31] De Jong, T. (2010). Cognitive load theory, educational research, and instructional design: some food for thought. Instructional Science. 38(2) pp 105-134.

[32] Dinkelbach, H., Vitay, J., Beuth, F., Hamker, F.H. (2012) Comparison of GPU- and CPU-implementations of mean-firing rate neural networks on parallel hardware. Network: Computation in Neural Systems., 23(4): 212-236.

[33] Dipietro, L., Sabatini, A. M., & Dario, P. (2008). Survey of glove-based systems and their applications. IEEE Transactions on systems, Man and Cybernetics, Part C: Applications and reviews, 38(4), 461-482.

[34] Ekman P., “Facial expression and emotion,” American Psychologist, vol. 48, no. 4, pp. 384-392, 1993.

[35] Ekman, P. & Oster, H. (1979). Facial Expressions of Emotion. Annual Review of Psychology, 30, 527-554.

[36] Ekman, P., Friesen, W., & Hager, J. (2002). Facial Action Coding System. Salt Lake City, UT: Research Nexus.

[37] El Ayadi, Moataz; Kamel, Mohamed S.; Karray, Fakhri (2011). Survey on speech emotion recognition: Features, classification schemes, and databases. Journal of Pattern Recognition, 44, S. 572-587.

[38] Eyben, F., Schuller, B., Rigoll G. (2012). Improving generalisation and robustness of acoustic affect recognition: Proceedings of the 14th ACM international conference on Multimodal interaction, pp. 517-522

[39] Eyben, F.; Wöllmer, M.; Schuller, B. „Opensmile: the munich versatile and fast open-source audio feature extractor.” In Proceedings of the international conference on Multimedia 2010 (pp. 1459-1462). ACM.

[40] Fan, R. E., Chang, K. W., Hsieh, C. J., Wang, X. R., & Lin, C. J. (2008). LIBLINEAR: A library for large linear classification. *Journal of machine learning research*, *9*(Aug), 1871-1874.

[41] Farid W.M., Mitropoulos F.J., Visualization and Scheduling of Non-functional Requirements for Agile Processes, In Proceedings of IEEE Southeastcon, pp. 1-8, 2013.

[42] Farinas, J., & Pellegrino, F. (2001). Automatic rhythm modeling for language identification. In *Seventh European Conference on Speech Communication and Technology*.

[43] Funke, J. 2010. Complex problem solving: a case for complex cognition? Cognitive Processing. 11, 2 (2010), 133–142.

[44] Findlatera L., McGrenere J., Beyond performance: Feature awareness in personalized interfaces, International Journal of Human-Computer Studies, vol. 68, iss. 3, pp. 121-137, 2010.

[45] Galy, E., Cariou, M., Mélan, C. „What is the relationship between mental workload factors and cognitive load types?“, *International Journal of Psychophysiology*, Bd. 83, Nr. 3, S. 269–275, März 2012.

[46] Gajos K., Everitt K., Tan D., Czerwinski M., Weld D., Predictability and Accuracy in Adaptive User Interfaces, In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 1271-1274, 2008.[DDM10] Dingsøyr T., Dybå T., Moe N., Agile Software Development: Current Research and Future Directions, Springer, 2010.

[47] Gudi, A., Tasli, H. E., den Uyl, T. M., & Maroulis, A. Deep Learning based FACS Action Unit Occurrence and Intensity Estimation

[48] Hall, J. Knapp, M. (2013): Nonverbal Communication. De Gruyter Mouton
[49] Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., & Witten, I. H. (2009). The WEKA data mining software: an update. *ACM SIGKDD explorations newsletter*, *11*(1), 10-18.
[50] Haq S., Asif M., Ali A., Jan T., Naveed Ahmad, Khan Y., (2016). Audio-visual emotion classification using filter and wrapper feature selection approaches: Sindh University Research Journal-SURJ (Science Series),vol. 47
[51] Herms, R. (2016). Prediction of Deception and Sincerity from Speech Using Automatic Phone Recognition-Based Features. In *INTERSPEECH* (pp. 2036-2040).
[52] Herms, R.; Ritter, M.; Wilhelm-Stein, T.; and Eibl, M., "Improving Spoken Document Retrieval by Un-supervised Language Model Adaptation Using Utterance-based Web Search". In 15th Annual Conference of the International Speech Communication Association (INTERSPEECH), 2014.
[53] Hofbauer, K., Petrik, S., & Hering, H. (2008, May). The ATCOSIM Corpus of Non-Prompted Clean Air Traffic Control Speech. In *LREC*.
[54] Huang, S.-T., Cho, Y.-P. & Lin, Y.-J. (2005). ADDIE Instruction Design and Cognitive Apprenticeship for Project-based Software Engineering Education in MIS. IEEE Computer Society - Proceedings of the 12th Asia-Pacific Software Engineering Conference (APSEC'05).
[55] Huang, W., „Handbook of Human Centric Visualization", Springer, 2014
[56] Hurter, C., R. Lesbordes, C. Letondal, J.-L. Vinot, and S. Conversy, "Strip'tic: Exploring augmented paper strips for air traffic controllers," in Proceedings of the International Working Conference on Advanced Visual Interfaces, ser. AVI'12. New York, NY, USA: ACM, 2012, pp. 225–23
[57] Iain Matthews, S. B. (2004). Active appearance models revisited. Kluwer Academic Publishers.
[58] ICAO, 2001. Aeronautical Telecommunications [online]. International Civil Aviation Organization [Zugriff am: 10.01.2018]. Verfügbar unter: https://www.icao.int/Meetings/anconf12/Document%20Archive/AN10_V2_cons%5B1%5D.pdf
[59] Imbert, J.-P. 2014. *Adaptation du design des visualisations de type supervisions pour optimiser la transmission des notifications classées par niveau d'intérêt*. Toulouse, ISAE.
[60] Imbert, J.-P., Hodgetts, H.M., Parise, R., Vachon, F., Dehais, F. and Tremblay, S. 2014. Attentional costs and failures in air traffic control notifications. *Ergonomics*. 57, 12 (2014), 1817–1832.
[61] Inoue, S., Furuta, K., Nakata, K., Kanno, T., Aoyama, H. & Brown, M. (2012). Cognitive process modelling of controllers in en route air traffic control. Ergonomics 55, 450-464.
[62] Jackson, P., & Haq, S. (2014). Surrey Audio-Visual Expressed Emotion(SAVEE) Database. *University of Surrey: Guildford, UK*.
[63] Josyula, D. P., Hughes, F. C., Vadali, H. Donahue, B. (2009). Modeling emotions for choosing between deliberation and action. Nature & Biologically Inspired Computing, NaBIC 2009, 782-787.
[64] Jurafsky, D., Shriberg, E., & Biasca, D. (1997). Switchboard-damsl labeling project coder's manu-al. Tech. Rep. 97-02.
[65] Jurafsky, D. (1997). Switchboard SWBD-DAMSL shallow-discourse-function annotation coders manual. *Institute of Cognitive Science Technical Report*.
[66] Juslin, P. N., & Scherer, K. R. (2005). Vocal expression of affect. In J. A. Harrigan, R. Rosenthal & K. R. Scherer (Eds.), The new handbook of methods on nonverbal behavior research (pp. 65–135). Oxford, UK: Oxford University Press.
[67] Kaber, D. B., Perry, C. M., Segall, N., McClernon, C. K. & Prinzel III, L. J., "Situation awareness implications of adaptive automation for information processing in an air traffic control-related task" International Journal of Industrial Ergonomics, 2006, 36, 447 – 462
[68] Kaess, M.; Johannsson, H.; Roberts, R.; Ila, V.; Leonard, J. & Dellaert, F. niSAM2: Incremental Smoothing and Mapping Using the Bayes Tree Intl. Journal of Robotics Research, 2 0 1 2, 31, 216−235
[69] Kalyuga, S. „Cognitive Load Theory: Implications for Affective Computing.", in *FLAIRS Conference*, 2011.
[70] Kanade, Takeo, Jeffrey F. Cohn, and Yingli Tian. "Comprehensive database for facial expression analysis." Automatic Face and Gesture Recognition, 2000. Proceedings. Fourth IEEE International Conference on. IEEE, 2000.
[71] Kludas, J., Bruno, E., Marchand-Maillet, S. „Information Fusion in Multimedia Information Retrieval", in *Adaptive Multimedia Retrieval: Retrieval, User, and Semantics*, Bd. 4918, N. Boujemaa, M. Detyniecki, und A. Nürnberger, Hrsg. Berlin, Heidelberg: Springer Berlin Heidelberg, 2008, S. 147–159.
[72] Koehn, P. (2005, September). Europarl: A parallel corpus for statistical machine translation. In *MT summit* (Vol. 5, pp. 79-86).

[73] Kumar, H., Agarwal, A., Dasgupta, R., Joshi, S., & Kumar, A. (2017). Dialogue Act Sequence Labeling using Hierarchical encoder with CRF. *arXiv preprint arXiv:1709.04250*.
[74] Kwang-Eun, K. and Kwee-Bo, S. (2010), Facial emotion recognition using a combining AAM with DBN, in Proc. Int. Conf. Control Autom. Syst., 1436–1439.
[75] Lafferty, J., McCallum, A., & Pereira, F. C. (2001). Conditional random fields: Probabilistic models for segmenting and labeling sequence data.
[76] Lamere, P., Kwok, P., Gouvea, E., Raj, B., Singh, R., Walker, W., & Wolf, P. (2003, April). The CMU SPHINX-4 speech recognition system. In *IEEE Intl. Conf. on Acoustics, Speech and Signal Processing (ICASSP 2003), Hong Kong* (Vol. 1, pp. 2-5).
[77] Lange, S. (2013). Faktorgraph-basierte Sensordatenfusion zur Anwendung auf einem Quadrocopter. Dissertation, TU Chemnitz.
[78] Langner, O., Dotsch, R., Bijlstra, G., Wigboldus, D.H.J., Hawk, S.T., & van Knippenberg, A. (2010). Presentation and validation of the Radboud Faces Database. *Cognition & Emotion, 24(8),* 1377—1388.
[79] Lavie, Talia; Meyer, Joachim (2010). Benefits and costs of adaptive user interfaces. In: International Journal Human-Computer Studies 68 (2010) 508–524.
[80] Le, Q. V. (2013). Building high-level features using large scale unsupervised learning. IEEE.
[81] Le, Q.V.; Ranzato, M.A.; Monga, R.; Devin, M.; Chen, K.; Greg S. Corrado, G.S., Jeff , D.; Ng, A.Y. (2012) Building High-level Features Using Large Scale Unsupervised Learning. In: International Conference on Machine Learning, 103.
[82] Lee, Chul Min; Narayanan, Shrikanth S.; Pieraccini, Roberto. „Combining acoustic and language information for emotion recognition.“ In: Proc. Of INTERSPEECH. 2002.
[83] Lehtonen T., Eloranta V.-P., Leppanen M., Isohanni E., Visualizations as a Basis for Agile Software Process Improvement, In Proceedings of 20th Asia-Pacific Software Engineering Conference, pp. 495-502, 2013.
[84] Liu, M., Li, S., Shan, S., and Chen, X. (2013), AU-aware Deep Networks for facial expression recognition, 10th IEEE Int. Conf. Work. Autom. Face Gesture Recognit. (IEEE), 1–6, 2013.
[85] Loft, S., Sanderson, P. Neal, A. Mooij, M. (2007). Modeling and Predicting Mental Workload in En Route Air. Traffic Control: Critical Review and Broader Implications. Human Factors: The Journal of the Human Factors and Ergonomics Society 49, 376-399.
[86] Luengo, I., Navas, E., Hernáez, I., Sánchez, J. (2005). Automatic Emotion Recognition using Prosodic Parameters. In: Proc. Interspeech 2005, 493-496.
[87] Maja Pantic, M. V. (2005). Web-based database for facial expression analysis. IEEE.
[88] Martin, O., Kotsia, I., Macq, B., & Pitas, I. (2006, April). The enterface'05 audio-visual emotion database. In *Data Engineering Workshops, 2006. Proceedings. 22nd International Conference on* (pp. 8-8). IEEE.
[89] Martins P., Caseiro R., Batista J., Generative face alignment through 2.5D active appearance models, Computer Vision and Image Understanding, vol. 117, no.3, pp. 250–268, 2013.
[90] Matthews L., Baker S., Active Appearance Models Revisited, International Journal of Computer Vision, vol. 60(2):135-164, 2004.
[91] Mengyi Liu, S. L. (2013). Au-aware deep networks for facial expression recognition. IEEE.
[92] Mikolov, T., Chen, K., Corr ado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. arXiv preprint arXiv:1301.3781.
[93] Miller, G., & Fellbaum, C. (1998). Wordnet: An electronic lexical database.
[94] Mitra, S. and Acharya, T. (2007), "Gesture Recognition: A Survey", IEEE Transactions on systems, Man and Cybernetics, Part C: Applications and reviews, 38(4), 311-317.
[95] „mpatacchiola/deepgaze", GitHub. [Online]. Verfügbar unter: https://github.com/mpatacchiola/deepgaze. [Zugegriffen: 13-Jan-2017].
[96] Mukherjee, S.S., Robertson, N.M., 2015. Deep Head Pose: Gaze-Direction Estimation in Multimodal Video. IEEE Transactions on Multimedia 17, 2094–2107. doi:10.1109/TMM.2015.2482819
[97] Müller, N. H. & Truschzinski, M. (2014). An Emotional Framework for a Smart Virtual Worker. In: Human-Computer Interaction. Advanced Interaction Modalities and Techniques Lecture Notes in Computer Science Volume 8511, 2014, pp 675-686.
[98] Müller, N. H. & Truschzinski, M. (2015). Analytical Steps for the calibration of an Emotional Framework – Pre-Test and Evaluation Procedures. In: Human-Computer Interaction: Design and Evaluation Lecture Notes in Computer Science Volume 9169, 2015, pp 512-519.
[99] Müller, N. H., Truschzinski, M., Fink, V., Schuster, J., Dinkelbach, H. Ü., Heft, W., Kronfeld, T., Rau, C. & Spitzhirn, M. (2014). The Smart Virtual Worker - Digitale Menschmodelle für die Simulation industrieller Arbeitsvorgänge. Technische Sicherheit Juli/August 2014.
[100] Mueller, S.C. (2011). The influence of emotion on cognitive control: Relevance for development and adolescent psychopathology. Frontiers in Psychology, 2(327).

[101] Navarrao, J. (2006): Menschen lesen. Mvg Verlag.
[102] Niegemann, H.M., Domagk, S., Hessel, S., Hein, A., Hupfer M. & Zobel, A. (2008): Kompendium multimediales Lernen. Heidelberg: Springer.
[103] OpenPose. https://github.com/CMU-Perceptual-Computing-Lab/openpose/blob/master/README.md
[104] Paas, F., van Gog, T. & Sweller, J. (2010). Cognitive Load Theory: New Conceptualizations, Specifications, and Integrated Research Perspectives. Educational Psychology Review. 22(2) pp 115-121.
[105] Palmer, E.M., T. C. Clausner, and P. J. Kellman, "Enhancing air traffic displays via perceptual cues," ACM Trans. Appl. Percept., vol. 5, no. 1, pp. 4:1–4:22, Jan. 2008
[106] Panayotov, V., Chen, G., Povey, D., & Khudanpur, S. (2015, April). Librispeech: an ASR corpus based on public domain audio books. In *Acoustics, Speech and Signal Processing (ICASSP), 2015 IEEE International Conference on* (pp. 5206-5210). IEEE.
[107] Pantic, Maja, et al. "Web-based database for facial expression analysis."Multimedia and Expo, 2005. ICME 2005. IEEE International Conference on. IEEE, 2005.
[108] Parasuraman, R., Mouloua, M., Hilburn, B. "Adaptive aiding and adaptive task allocation enhance human-machine interaction" Automation technology and human performance: Current research and trends, 1999, 119-123
[109] Perin C., Vuillemot R., Fekete J.-D., SoccerStories: A Kick-off for Visual Soccer Analysis, IEEE Transactions on Visualization and Computer Graphics, vol. 19, no. 12, pp. 2506-2515, 2013.
[110] Peterson, L., Bailey, L., & Willems, B. (2001). Controller-to-controller communication and coordination taxonomy (C^4T). (DOT/FAA/AM-01/19). Department of Transportation, Federal Aviation Administration, Office of Aerospace Medicine, Washington, D.C.
[111] Pfeiffer, L., Sims, T. and Rosenthal, P. 2017. Visualizing Workload and Emotion Data in Air Traffic Control - An Approach Informed by the Supervisors Decision Making Process. *Proceedings of the International Conference on Advances in Computer-Human Interactions* (2017), 81–87.
[112] Pfeiffer, L., Müller, N. H. & Rosenthal, P. (2015). A Survey of Visual and Interactive Methods for Air Traffic Control Data. Proceedings of the International Conference on Information Visualisation, pp. 574–577.
[113] Pfeiffer, L., Müller, N. H., Valtin, G., Truschzinski, M.; Ohler, P. & Rosenthal, P. (2015). Emotionsmodell für zukünftige Mensch-Technik-Schnittstellen zur Unterstützung von Centerlotsen. DGLR-Bericht 2015-01.
[114] Pfeiffer, L., Valtin, G., Müller, N. H. & Rosenthal, P. (2015). Aircraft in Your Head: How Air Traffic Controller Mentally Organize Air Traffic. Pascal Lorenz, Christian Bourret (Ed.): Proceedings of HUSO, the International Conference on Human and Social Analytics, pp. 19–24, IARIA (Best Paper Award)
[115] Pfeiffer, L., Valtin, G., Müller, N.H. and Rosenthal, P. 2016. The Mental Organization of Air Traffic and its Implications to an Emotion Sensitive Assistance System. International Journal on Advances in Life Sciences. 8, (2016).
[116] Peng, X., Huang, J., Hu, Q., Zhang, S., Metaxas, D.N., 2015. Three-dimensional head pose estimation in-the-wild, in: 2015 11th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG). Presented at the 2015 11th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG), pp. 1–6. doi:10.1109/FG.2015.7163109
[117] Phung, T., Winikoff, M. & Padgham, L. (2005). Learning within the BDI Framework: An Empirical Analysis. In: Knowledge-Based Intelligent Information and Engineering Systems Lecture Notes in Computer Science Volume 3683, 2005, pp 282-288
[118] Pittermann, Johannes; Pittermann, Angela; Minker, Wolfgang. „Emotion recognition and adaptation in spoken dialogue systems." In: International Journal of Speech Technology, 2010, 13. Jg., Nr. 1, S. 49-60.
[119] Quesada, J., Kintsch, W. and Gomez, E. 2005. Complex problem-solving: a field in search of a definition? *Theoretical Issues in Ergonomics Science*. 6, 1 (2005), 5–33.
[120] Radeck-Arneth, S., Milde, B., Lange, A., Gouvêa, E., Radomski, S., Mühlhäuser, M., & Biemann, C. (2015, September). Open source german distant speech recognition: Corpus and acoustic model. In *International Conference on Text, Speech, and Dialogue* (pp. 480-488). Springer, Cham.
[121] Rao, A. S. & Georgeff, M. P. (1995). BDI Agents: From Theory to Practice. Proceedings of the First International Conference on Multiagent Systems. ICMAS-95, pp 312-319
[122] Rau C., Brunnett G.: Anatomically correct adaption of kinematic skeletons to virtual humans , Proceedings of the International Conference on Computer Graphics Theory and Applications (GRAPP), pp. 341-346

[123] Rau C., Brunnett G.: GPU-accelerated Real-time Markerless Human Motion Capture, Proceedings of the 8th International Conference on Computer Graphics Theory and Applications (GRAPP), 2013, pp. 397-401

[124] Redding, R. E. & Seamster, T. L. (1996). Cognitive Task Analysis of Air Traffic Control Instruction to Identify Rule-Based Measures of Student Simulator Performance. Proceedings of the Human Factors and Ergonomics Society Annual Meeting. 40 (4) pp 269-273.

[125] Reger, K., & Ohler, P. (1999). Emotional cooperating agents and group formation. A system analysis of role-play amoung children. Modelling and Simulation - A Tool for the Next Millenium, 13th European Simulation Multiconference (Proceedings). Erlangen, San Diego: SCS Publication.

[126] Reichel, U.D., Kisler, T. (2014). Language-independent grapheme-phoneme conversion and word stress assignment as a web service. In: Hoffmann, R. (Ed.): Elektronische Sprachverarbeitung. Studientexte zur Sprachkommunikation 71, pp 42-49, TUDpress, Dresden.

[127] Ritter, M., Herms, R., Manthey, R., Eibl, M., "Ein ganzheitlicher Ansatz zur Digitalisierung und Extraktion von Metadaten in Videoarchiven" In: Proceedings des 13. Internationalen Symposiums für Informationswissenschaft (ISI 2013), S.362-371.

[128] Ren, S., He, K., Girshick, R., & Sun, J. (2015). Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *arXiv:1506.01497 [cs]*. Abgerufen von http://arxiv.org/abs/1506.01497

[129] Rosch, J. L., & Vogel-Walcutt, J. J. (2012). A review of eye-tracking applications as tools for training. Cognition, Technology & Work. August 2013, Volume 15, Issue 3, 313-327.

[130] Rosenthal, P., Pfeiffer, L., Müller, N. H. & Ohler, P. (2013). VisRuption: Intuitive and Efficient Visualization of Temporal Airline Disruption Data. Computer Graphics Forum, vol. 32, iss. 3, pp. 81–90, 2013.

[131] Rosenthal, P., Pfeiffer, L., Müller, N.H. and Valtin, G. 2017. The Long Way to Intuitive Visual Analysis of Air Traffic Control Data. *Visual Analytics for Management: Translational Science and Applications in Practice.* E. Bendoly and S. Clark, eds. Routledge. 138–148.

[132] Rosman B., Ramamoorthy S., Mahmud H., Kohli P., On User Behaviour Adaptation Under Interface Change, In Proceedings of the 19th International Conference on Intelligent User Interfaces, pp. 273-278, 2014.

[133] Salamah, S. & Brunnett, G., Hierarchical Method for Segmentation by Classification of Motion Capture Data, To appear in "Virtual Realities", Proc. Dagstuhl Seminar 2013 on Virtual Realities, LNCS, Springer

[134] Salden, R. J. C. M., Paas, F., Jeroen, J. G. & van Merrienboer, J. (2006). Personalised adaptive task selection in air traffic control: Effects on training efficiency and transfer. Learning and Instruction. 16(4). Pp 350-362.

[135] Schaab, B. "The Influence of Ascending and Descending Levels of Workload on Performance" Automation Technology and Human Performance: Current research and trends, 1999, 218-220

[136] Schmidt, Thomas & Schütte, Wilfried (2010): FOLKER: An Annotation Tool for Efficient Transcription of Natural, Multi-party Interaction. in: Proceedings of LREC 2010.

[137] Schuller, B., Steidl, S., Batliner, A., Burkhardt, F., Devillers, L., Müller, C. A., & Narayanan, S. S. (2010, September). The INTERSPEECH 2010 paralinguistic challenge. In INTERSPEECH (pp. 2794-2797).

[138] Schuller, B.; Rigoll, G.; Lang, M. „Speech emotion recognition combining acoustic features and linguistic information in a hybrid support vector machine-belief network architecture." In: Acoustics, Speech, and Signal Processing, 2004. Pro-ceedings.(ICASSP'04). IEEE International Conference on Vol. 1, S. I-577.

[139] Schuller, B., Vlasenko, B., Eyben, F., Rigoll, G., & Wendemuth, A. (2009, November). Acoustic emotion recognition: A benchmark comparison of performances. In *Automatic Speech Recognition & Understanding, 2009. ASRU 2009. IEEE Workshop on* (pp. 552-557). IEEE.

[140] Schuller, B., Vlasenko, B., Eyben, F., Wollmer, M., Stuhlsatz, A., Wendemuth, A., & Rigoll, G. (2010). Cross-corpus acoustic emotion recognition: Variances and strategies. *IEEE Transactions on Affective Computing*, *1*(2), 119-131.

[141] Schultz, T., & Waibel, A. (2001). Language-independent and language-adaptive acoustic modeling for speech recognition. *Speech Communication*, *35*(1-2), 31-51.

[142] Schulz, H.-J.; Nocke, T.; Heitzler, M.; Schumann, H., „A Design Space of Visualization Tasks," Visualization and Computer Graphics, IEEE Transactions on , vol.19, no.12, pp.2366-2375, Dec. 2013.

[143] Sears, A. and Jacko, J.A. eds. 2007. *The Human-Computer Interaction Handbook: Fundamentals, Evolving Technologies, and Emerging Applications*. Lawrence Erlbaum Assoc Inc.

[144] Selting, M., Auer, P., Barth-Weingarten, D., Bergmann, J. R., Bergmann, P., Birkner, K., ... & Hartung, M. (2009). Gesprächsanalytisches transkriptionssystem 2 (GAT 2). Gesprächsforschung: Online-Zeitschrift zur verbalen Interaktion

[145] Shingade, A., Ghotkar, A. (2014): Animation of 3D Human Model Using Markerless Motion Capture Applied To Sports. International Journal of Computer Graphics & Animation (IJCGA) Vol.4, No.1, January 2014
[146] Shotton, J., Fitzgibbon, A., Cook, M., Sharp, T., Finocchio, M., Moore, R., Kipman, A. and Blake, A. (2011). Real-time human pose recognition in parts from a single depth image. CVPR 2011
[147] Simon, T., Joo, H., Matthews, I., and Sheikh, Y. *Hand Keypoint Detection in Single Images Using Multiview Bootstrapping* (In CVPR 2017)
[148] Smallman, H., M. St.John, H. Oonk, and M. Cowen, "Information availability in 2d and 3d displays,"Computer Graphicsand Applications, IEEE, vol. 21, no. 5, pp. 51–57, Sep 2001
[149] Soraji, Y., Furuta, K., Kanno, T., Aoyama, H., Inoue, S., Karikawa, D., et al. (2010). Cognitive Model of Team Cooperation in En-route Air Traffic Control. Cognition, Technology & Work, 14(2), 93–105. doi:10.1007/s10111-010-0168-x.
[150] Stolcke, A. (2002). SRILM-an extensible language modeling toolkit. In *Seventh international conference on spoken language processing*.
[151] Tavanti, M., H.-H. Le, and N.-T. Dang, "Three-dimensional stereoscopic visualization for air traffic control interfaces: a preliminary study," in Digital Avionics Systems Conference, 2003. DASC '03. The 22nd, vol. 1, Oct 2003, pp. 5.A.1–5.1–7 vol.1.
[152] Teichmann, M., Wiltschut, J., Hamker, F.H. (2012) Learning invariance from natural images in-spired by observations in the primary visual cortex. Neural Computation, 24(5): 1271-1296.
[153] Team, R. C. (2013). R: A language and environment for statistical computing.
[154] Toshev, A., Szegedy, C., 2014. DeepPose: Human Pose Estimation via Deep Neural Networks, in: 2014 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Presented at the 2014 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1653–1660. doi:10.1109/CVPR.2014.214
[155] Truschzinski, M. and Müller, N. (2014). An emotional model for social robots: late-breaking report. In Proceedings of the 2014 ACM/IEEE international conference on Human-robot interaction (HRI '14). ACM, New York, NY, USA, 304-305.
[156] Truschzinski, M., Müller, N. H., Dinkelbach, H.-U., Protzel, P., Hamker, F. & Ohler, P. (2014). Deducing human emotions by robots: Computing basic non-verbal expressions of performed actions during a work task. ICIS 2014, Auckland.
[157] Truschzinski, M., Pfeiffer, L. and Valtin, G. 2017. ELSI-Aspekte in Forschungsverbünden. Mensch und Computer 2017 - Workshopband (Regensburg, 2017).
[158] Twitter4J - A Java library for the Twitter API. [Online]. Verfügbar unter: http://twitter4j.org/en/index.html. [Zugegriffen: 17-Juni-2015].
[159] Vasquez-Correa, J. C., Arias-Vergara, T., Orozco-Arroyave, J. R., Vargas-Bonilla, J. F., Noeth E., (2016). Wavelet-Based Time-Frequency Representations for Automatic Recognition of Emotions from Speech: Speech Communication; 12. ITG Symposium; VDE; pp. 1-5
[160] Wells, J. C. (1997). SAMPA computer readable phonetic alphabet. *Handbook of standards and resources for spoken language systems*, *4*.
[161] Wittenburg, P., Brugman, H., Russel, A., Klassmann, A., Sloetjes, H. (2006).
ELAN: a Professional Framework for Multimodality Research.
In: Proceedings of LREC 2006, Fifth International Conference on Language Resources and Evaluation.
[162] Tufte, E. 1990. *Envisioning Information.* Graphics Press.
[163] Ververides, Dimitrios; Kotropoulos, Constantine (2003). A Review of Emotional Speech Databases. In: Proc. 9th Panhellenic Conference on Informatics, S. 560-570
[164] Vogt, J., Hagemann, T. & Kastner, M. (2006). The impact of workload on heart rate and blood pressure in en-route and tower air traffic control. Journal of Psychophysiology, 20, 297-314.
[165] Vinot, J.-L., C. Letondal, R. Lesbordes, S. Chatty, S. Conversy, C. Hurter, "Tangible augmented reality for air traffic control," interactions, vol. 21, no. 4, pp. 54–57, Jul. 2014
[166] Vortac, O. U., Edwards, M. B., Fuller, D. K. & Manning, C. A. (1993). Automation and cognition in air traffic control: An empirical investigation. Applied Cognitive Psychology 7(7) pp 631-651.
[167] Wagner, J., Lingenfelser, F., André, E., Kim J. (2011). Exploring Fusion Methods for Multimodal Emotion Recognition with Missing Data. IEEE Transactions on Affective Computing.
[168] Wang, J. M., Fleet, D. J., Hertzmann, A., Gaussian Process Dynamical Models for Human Motion. IEEE Transactions on Pattern Analysis and Machine Intelligence, 30(2):283-296, 2008.
[169] Ware, C., „Information Visualization: Perception for Design", Morgan Kaufmann, 2013
[170] Wei, S.-E., Ramakrishna, V., Kanade, T., Sheikh, Y., „Convolutional Pose Machines", ArXiv160200134 Cs, Jan. 2016.

[171] Weinland, D., Ronfard, R. and Boyer, E. (2011). A survey of vision-based methods for action representation, segmentation and recognition. Computer Vision and Image Understanding, 115(2):224–241.
[172] Wilhelm-Stein, T., Herms, R., Ritter, M., Eibl, M., “Improving Transcript-Based Video Retrieval Using Unsupervised Language Model Adaptation”. In: Information Access Evaluation. Multilinguality, Multimodality, and Interaction. Springer International Publishing, 2014. S. 110-115
[173] Wimmer, M. u.a. (2008) LOW-LEVEL FUSION OF AUDIO AND VIDEO FEATURE FOR MULTI-MODAL EMOTION RECOGNITION. VISAPP
[174] Witzke, D., L. Pescheck, and M. Miosgae, “Tess - Fluglotsenarbeitsplatz,” Bachelorthesis, Hochschule für Gestaltung Schwäbisch Gmünd, 2013.
[175] Wong, BL William, Simone Rozzi, Alessandro Boccalatte, Stephen Gaukrodger, Paola Amaldi, Bob Fields, Martin Loomes, and Peter Martin. "3D-in-2D Displays for ATC." In *6th EUROCONTROL Innovative Research Workshop*, pp. 42-62. 2007
[176] Zhou, X., Zhu, M., Leonardos, S., Derpanis, K., Daniilidis, K., 2015. Sparseness Meets Deepness: 3D Human Pose Estimation from Monocular Video. arXiv:1511.09439 [cs].
[177] Zillmann, D., Johnson, R.C. and Day, K. D. Attribution of apparent arousal and proficiency of recovery from sympathetic activation affecting excitation transfer to aggressive behavior. Journal of Experimental Social Psychology, 10(6):503–515, 1974.
[178] Krizhevsky, I. Sutskever, G. E. Hinton. Imagenet classification with deep convolutional neural networks, in: Advances in neural information processing systems, pp. 1097-1105.
[179] Y. Tang. Deep learning using support vector machines. CoRR, abs/1306.0239
[180] M. F. Valstar, T. Almaev, J. M. Girard, G. McKeown, M. Mehu, L. Yin, M. Pantic, J. F. Cohn. Fera 2015-second facial expression recognition and analysis challenge, in: Automatic Face and Gesture Recognition (FG), 2015 11th IEEE International Conference and Workshops on, volume 6, IEEE, pp. 1-8.
[181] Zhang, Q. 2012. *Theoretical Review on a 3D based Air Traffic Control System*. Uppsala University.
[182] Kirwan, B., Donohoe, L., Atkinson, T., MacKendrick, H., Lamoureux, T. and Phillips, A. 1998. Getting the picture-Investigating the mental picture of the air traffic controller. *Contemporary ergonomics 1998* (1998), 404–408.
[183] Niessen, C. and Eyferth, K. 2001. A model of the air traffic controller’s picture. *Safety Science*. 37, 2–3 (2001), 187–202.
[184] Nunes, A. and Mogford, R.H. 2003. Identifying Controller Strategies that Support the ’Picture. *PROCEEDINGS of the HUMAN FACTORS AND ERGONOMICS SOCIETY 47th ANNUAL MEETING*. 47, 1 (2003), 71–75.
[185] Shorrock, S.T. and Isaac, A. 2010. Mental Imagery in Air Traffic Control. *International Journal of Aviation Psychology*. 20, 4 (2010), 309–324.
[186] Smallman, H.S., St.John, M., Oonk, H.M. and Cowen, M.B. 2001. Information availability in 2D and 3D displays. *Computer Graphics and Applications, IEEE*. 21, 5 (Sep. 2001), 51–57.
[187] Tavanti, M. and Cooper, M. 2009. Looking for the 3D Picture: The Spatio-temporal Realm of Student Controllers. *Human Centered Design*. M. Kurosu, ed. Springer Berlin Heidelberg. 1070–1079.
[188] Zühlke, D. 2012. Nutzergerechte Entwicklung von Mensch-Maschine-Systemen: Useware-Engineering für technische Systeme. Springer Berlin Heidelberg. 5–34.

Chemnitzer Informatik-Berichte

In der Reihe der Chemnitzer Informatik-Berichte sind folgende Berichte erschienen:

CSR-10-01 Maximilian Eibl, Jens Kürsten, Robert Knauf, Marc Ritter , Workshop Audiovisuelle Medien, Mai 2010, Chemnitz

CSR-10-02 Thomas Reichel, Gudula Rünger, Daniel Steger, Haibin Xu, IT-Unterstützung zur energiesensitiven Produktentwicklung,
Juli 2010, Chemnitz

CSR-10-03 Björn Kre llner, Thomas Reichel, Gudula Rünger, Marvin Ferber, Sascha Hunold, Thomas Rauber, Jürgen Berndt, Ingo Nobbers, Transformation monolithischer Business-Softwaresysteme in verteilte, workflowbasierte Client-Server-Architekturen, Juli 2010, Chemnitz

CSR-10-04 Björn Krellner, Gudula Rünger, Daniel Steger, Anforderungen an ein Datenmodell für energiesensitive Prozessketten von Powertrain-Komponenten, Juli 2010, Chemnitz

CSR-11-01 David Brunner, Guido Brunnett, Closing feature regions, März 2011, Chemnitz

CSR-11-02 Tom Kühnert, David Brunner, Guido Brunnett, Betrachtungen zur Skelettextraktion umformtechnischer Bauteile, März 2011, Chemnitz

CSR-11-03 Uranchimeg Tudevdagva, Wolfram Hardt, A new evaluation model for eLearning programs, Dezember 2011, Chemnitz

CSR-12-01 Studentensymposium Informatik Chemnitz 2012, Tagungsband zum 1. Studentensymposium Chemnitz vom 4. Juli 2012, Juni 2012, Chemnitz

CSR-12-02 Tom Kühnert, Stephan Rusdorf, Guido Brunnett, Technischer Bericht zum virtuellen 3D-Stiefeldesign, Juli 2012, Chemnitz

CSR-12-03 René Bergelt, Matthias Vodel, Wolfram Hardt, Generische Datenerfassung und Aufbereitung im Kontext verteilter, heterogener Sensor-Aktor-Systeme, August 2012, Chemnitz

CSR-12-04 Arne Berger, Maximilian Eibl, Stephan Heinich, Robert Knauf, Jens Kürsten, Albrecht Kurze, Markus Rickert, Marc Ritter , Schlussbericht zum InnoProfile Forschungsvorhaben sachsMedia - Cooperative Producing, Storag, Retrieval and Distribution of Audiovisual Media (FKZ: 03IP608), September 2012, Chemnitz

CSR-12-05 Anke Tallig, Grenzgänger - Roboter als Mittler zwischen der virtuellen und realen sozialen Welt, Oktober 2012, Chemnitz

Chemnitzer Informatik-Berichte

CSR-13-01 Navchaa Tserendorj, Uranchimeg Tudevdagva, Ariane Heller, Grenzgänger - Integration of Learning Management System into University-level Teaching and Learning, Januar 2013, Chemnitz

CSR-13-02 Thomas Reichel, Gudula Rünger, Multi-Criteria Decision Support for Manufacturing Process Chains, März 2013, Chemnitz

CSR-13-03 Haibin Xu,Thomas Reichel, Gudula Rünger, Michael Schwind, Softwaretechnische Verknüpfung der interaktiven Softwareplattform Energy Navigator und der Virtual Reality Control Platform, Juli 2013, Chemnitz

CSR-13-04 International Summerworkshop Computer Science 2013, Proceedings of International Summerworkshop 17.7. - 19.7.2013, Juli 2013, Chemnitz

CSR-13-05 Jens Lang, Gudula Rünger, Paul Stöcker, Dynamische Simulationskopplung von Simulink-Modellen durch einen Functional-Mock-up-Interface- Exportfilter, August 2013, Chemnitz

CSR-14-01 International Summerschool Computer Science 2014, Proceedings of Summerschool 7.7.-13.7.2014, Juni 2014, Chemnitz

CSR-15-01 Arne Berger, Maximilian Eibl, Stephan Heinich, Robert Herms, Stefan Kahl, Jens Kürsten, Albrecht Kurze, Robert Manthey, Markus Rickert, Marc Ritter, ValidAX - Validierung der Frameworks AMOPA und XTRIEVAL, Januar 2015, Chemnitz

CSR-15-02 Maximilian Speicher, What is Usability? A Characterization based on ISO 9241-11 and ISO/IEC 25010, Januar 2015, Chemnitz

CSR-16-01 Maxim Bakaev, Martin Gaedke, Sebastian Heil, Kansei Engineering Experimental Research with University Websites, April 2016, Chemnitz

CSR-18-01 Jan-Philipp Heinrich, Carsten Neise, Andreas Müller, Ähnlichkeitsmessung von ausgewählten Datentypen in Datenbanksystemen zur Berechnung des Grades der Anonymisierung, Februar 2018, Chemnitz

CSR-18-02 Liang Zhang, Guido Brunnett, Efficient Dynamic Alignment of Motions, Februar 2018, Chemnitz

CSR-18-03 Guido Brunnett, Maximilian Eibl, Fred Hamker, Peter Ohler, Peter Protzel, StayCentered - Methodenbasis eines Assistenzsystems für Centerlotsen (MACeLot) Schlussbericht, November 2018, Chemnitz

Chemnitzer Informatik-Berichte

ISSN 0947-5125

Herausgeber: Fakultät für Informatik, TU Chemnitz
Straße der Nationen 62, D-09111 Chemnitz