

Universität der Bundeswehr München
Fakultät für Luft- und Raumfahrttechnik
Institut für Flugsysteme

Bewertung psychophysiologischer Mess- Sensoriken als Basis für nutzeradaptive Automation

Dennis Mund M.Sc.

Vollständiger Abdruck der bei der
Fakultät für Luft- und Raumfahrttechnik
der Universität der Bundeswehr München
zur Erlangung des akademischen Grades eines

Doktor-Ingenieurs (Dr.-Ing.)

genehmigten Dissertation.

1. Berichterstatter: Univ.-Prof. Dr.-Ing. Axel Schulte
2. Berichterstatter: Univ. -Prof. Dr.-Ing. Nele Rußwinkel

Diese Dissertation wurde am 22.06.2022 bei der Universität der Bundeswehr München,
85577 Neubiberg, eingereicht und durch die Fakultät für Luft- und Raumfahrttechnik am
23.11.2022 angenommen.

Tag der mündlichen Prüfung: 09.12.2022

Für meinen Großvater

Vorwort

Das Erstellen dieser Dissertation war ein weitaus größeres und anstrengenderes Unterfangen als erwartet. Gleichzeitig war die Arbeit allerdings jederzeit voller interessanter Probleme und damit verbundenen ausgiebigen Diskussionen.

An dieser Stelle möchte ich mich bei allen Leuten bedanken, die einen Teil dazu beigetragen haben, dass es zum Abschluss dieser Arbeit kommen konnte.

Allen voran habe ich hierbei meiner Freundin Margit zu danken, die die Stelle als wissenschaftlicher Mitarbeiter für mich gefunden hat und mir die gesamte Zeit über den Rücken gestärkt hat, vor allem wenn es mal nicht so lief.

Zusätzlich gebührt auch meinem Doktorvater, Herrn Univ.-Prof. Dr.-Ing Axel Schulte, ein besonderer Dank, dafür dass er die Stelle erst einmal ermöglicht hat. Gleichzeitig ist auch die Art und Weise wie er die Mitarbeiter dazu ermutigt ihr eigenes Thema zu finden und auszugestalten etwas was ich mir nicht besser hätte vorstellen können.

Weiterhin Möchte ich Frau Univ.-Prof. Dr.-Ing. Nele Rußwinkel für die Übernahme des Koreferats danken.

Danke auch an Diana Donath, dafür dass sie jederzeit ein offenes Ohr für jeden Mitarbeiter hat und im Hintergrund die Fäden zieht, sodass jeder sich auf seine Arbeit konzentrieren kann.

Natürlich gebührt ein großer Dank auch allen Kollegen, die mich in dieser Zeit begleitet haben und somit maßgeblich beteiligt waren. Hierbei sind insbesondere die Mitarbeiter des Jet-Simulators zu erwähnen. Allen voran Felix Heilemann und Florian Reich, mit denen diese wundersame Reise losging. Außerdem aber auch Thomas Leichtle, Sebastian Linder und Simon Schwerd, von denen ich viel gelernt habe. Reduzierten Dank möchte ich noch an Karl Tschurtschenthaler aussprechen, dafür dass er viele meiner Tätigkeiten übernommen und dadurch dafür gesorgt hat, dass ich in Ruhe schreiben konnte.

Zu guter Letzt möchte ich noch meiner Familie dafür danken, dass sie mich zu dem Menschen gemacht haben, der ich bin.

Vielen Dank!

Kurzfassung

Diese Dissertation befasst sich mit der Ermöglichung von automatisierter Unterstützung, die sich an die **Beanspruchung** von Flugzeugführern zukünftiger Teams aus bemannten und unbemannten Flugzeugen **adaptiert**. Die Grundvoraussetzung dafür bietet eine Echtzeitbestimmung der Beanspruchung zur Identifikation von Situationen mit erhöhtem Unterstützungsbedarf. Im Rahmen der Arbeit wurden dazu unterschiedliche Methoden der Operationalisierung von Beanspruchung experimentell untersucht und daraufhin auf ihre Eignung für **adaptive Automation** bewertet.

Vergangene Untersuchungen von adaptiver Automation trennen häufig die Analyse der Beanspruchung oder anderer Beschreibungen des mentalen Zustands von der Auswahl einer adäquaten Unterstützungsstrategie. Im Gegensatz dazu soll in dieser Arbeit betrachtet werden, inwiefern bereits die Bestimmung der Beanspruchung Anhaltspunkte für das Aussehen einer Adaption der Automation geben kann.

Eine zusätzliche Kernkomponente der Arbeit stellt die **Entwicklung eines Funktionsprototypen** dar, der eine realistische Simulation eines komplexen reellen Aufgabenspektrums abbildet. Nur in solchen Simulationen können die Wechselwirkungen von adaptiver Automation mit dem Menschen in closed-loop Experimenten untersucht werden.

Innerhalb des Funktionsprototypen werden unterschiedliche Arten der Beanspruchungsbestimmung **experimentell untersucht** und die Erkenntnisse über die Methoden anhand von **Bewertungskriterien** zusammengefasst. Abschließend wird aufbauend auf der vorangegangenen Bewertung ein **Konzept** für den Einsatz von adaptiver Automation für den **Führer eines bemannt-unbemannten Teams** vorgestellt, welches die Vorteile der jeweiligen Messmethoden vereint.

Abstract

This dissertation addresses the enabling of automated support that adapts to the workload experienced by pilots of future manned- unmanned aircraft teams. The basic prerequisite for this is a real-time determination of the workload to identify situations with an increased need for support. For this purpose, different methods of operationalizing workload were investigated experimentally and then evaluated for their suitability for adaptive automation.

Past studies of adaptive automation often separate the analysis of workload or other mental state descriptions from the selection of an adequate support strategy. In contrast, this work will consider the extent to which even the determination of strain can provide clues as to the appearance of an adaptation of automation.

An additional core component of the work is the development of a functional prototype, which represents a realistic simulation of a complex real task spectrum. Only in such simulations can the interactions of adaptive automation with humans be investigated in closed-loop experiments.

Within the functional prototype, different types of workload determination are experimentally investigated and the findings about the methods are summarized based on evaluation criteria. Finally, based on the previous evaluation, a concept for the use of adaptive automation for the leader of a manned-unmanned team is presented, which combines the advantages of the respective measurement method

Inhalt

1. Einführung	3
1.1 Forschungsfrage.....	4
1.2 Überblick über die Arbeit	5
2. Mensch-Maschine-Systeme	7
2.1 Der Einsatz von Automation	7
2.2 Der Mensch als überwachende Instanz.....	8
2.3 Klassische Mensch-Maschine Aufgabenteilung.....	11
2.3.1 ‚Left-over‘ Zuweisung	11
2.3.2 Kompensationsbasierte Zuweisung	12
2.3.3 Komplementäre Zuweisung	12
2.4 Probleme beim Einsatz von Automation	14
2.4.1 Skill degradation	14
2.4.2 Out-of-the-loop performace	14
2.4.3 Trust in Automation	15
2.4.4 Schlechte Ressourcenauslastung des Menschen.....	17
2.5 Adaptive Automation.....	18
2.5.1 Adaptionmöglichkeiten	18
2.5.2 Auslöser für Adaptionen	21
2.6 Konsolidierung	31
3. Operationalisierung menschlicher Beanspruchung	32
3.1 Modellierung von Beanspruchung.....	32
3.1.1 Kognitive Architekturen	33
3.1.2 VACP.....	35
3.1.3 Single Resource Theory	37
3.1.4 Multiple Resource Theory	39
3.1.5 Cognitive Reliability and Error Analysis Method	41
3.1.6 Cognitive Task Load Modell	44
3.1.7 NASA-TLX.....	46
3.2 Messung von Beanspruchungsindikatoren	46
3.2.1 Kontextfreie Ermittlungsmethoden.....	47
3.2.2 Kontextbehaftete Ermittlungsmethoden	53
3.3 Konsolidierung	58

4. Untersuchung psychophysiologischer Messmethoden	60
4.1 Bewertungskriterien	60
4.1.1 Robustheit/Selektivität.....	60
4.1.2 Sensitivität	61
4.1.3 Diagnostizität.....	61
4.1.4 Individualität.....	61
4.1.5 Generalisierbarkeit.....	61
4.1.6 Direkte Anwendbarkeit.....	62
4.2 Manned-unmanned Teaming Kampfjetsimulator.....	62
4.2.1 Missionsszenario	63
4.2.2 Mensch-Maschine Schnittstelle	64
4.2.3 Simulatorarchitektur	66
4.2.4 Blickbewegungsmesssystem	70
4.2.5 Aktivitätsbestimmung.....	75
4.3 Kontextfreie Ermittlungsmethoden	80
4.3.1 Pupillenbasiertes Experiment	80
4.3.2 Elektrokardiografbasiertes Experiment	89
4.4 Kontextbehaftete Ermittlungsmethoden.....	102
4.4.1 Verhaltensbasiertes Experiment	102
4.4.2 Modellbasiertes Messverfahren.....	111
5. Diskussion.....	122
6. Zusammenfassung und Ausblick	125
Abkürzungsverzeichnis.....	127
7. Literaturverzeichnis	129

1. Einführung

Das 21. Jahrhundert ist geprägt von Fortschritten im Bereich der Automatisierung. Industrieroboter übernehmen viele Tätigkeiten, die vormals vom Menschen durchgeführt wurden. Zusätzlich wird unter dem Namen ‚Industrie 4.0‘ die Zusammenarbeit von Menschen und Robotern vorangetrieben. Dies geschieht beispielsweise, indem die Zäune, in denen Roboter aus Sicherheitsgründen früher agieren mussten, entfernt werden und Aufgaben gemeinsam durchgeführt werden können. In Projekten wie ‚*Franka.Emika*‘ (Haddadin et al. 2022), oder den LWR-Roboter der Firma Kuka (Schreiber et al. 2010) wird dabei versucht die Roboter sicher zu machen, indem sie kleinste Kräfte sensieren und somit potentiellen Gefahrensituationen frühzeitig entgegenwirken können.

Parallel dazu sorgen Fortschritte im ‚Autonomen Fahren‘ dafür, dass neben Industrieanwendungen auch Privatanwender immer mehr Kontakt mit hochautomatisierten Systemen haben. Der Übergang zum autonomen Fahren führt dabei dazu, dass zusätzliche Assistenzfunktionen in die Fahrzeuge eingerüstet werden, die im Falle von Teslas ‚*Full Self-Driving*‘ (Xu et al. 2022) den Nutzer bereits in eine reine Überwachungsrolle versetzen.

Einen weiteren Anwendungsfall für die Kooperation von Menschen mit einer entsprechend gearteten Automation stellt das sogenannte ‚Human-Autonomy Teaming‘ dar, bei dem menschliche Operateure mit hochautomatisierten, in der Regel KI-basierten Agenten ein gemeinsames Ziel erreichen sollen. Eine Anwendung dieser Kooperation bildet dabei das Manned-Unmanned-Teaming (MUM-T). Dieser militärisch geprägte Begriff, beschreibt das Zusammenwirken von bemannten und unbemannten Militärluftfahrzeugen (vorwiegend Helikopter und Jet-Flugzeuge) um gemeinsam eine militärische Mission durchzuführen, wie beispielsweise die Bekämpfung eines Hochwertziels (Kampfflugzeug) oder die Bergung verbündeter Kräfte (Helikopter).

Diese Anwendungsfälle haben als Gemeinsamkeit, dass die jeweiligen Operateure die automatisierten Systeme überwachen müssen und in Fehlerfällen eingreifen, um potenzielle Gefahren zu verhindern. Gleichzeitig bedingen die dazugehörigen Anwendungsfälle komplexe Arbeitsabläufe der Automation, bei denen es für den Operateur schwierig ist diese zu verstehen und Fehlerfälle zuverlässig zu erkennen. Beispielsweise erlaubt es die höhere Reaktionsgeschwindigkeit von automatisierten Automobilen, dass solche Fahrzeuge einen geringeren Sicherheitsabstand zum vorherfahrenden Fahrzeug einhalten. Dies führt dazu, dass die Testfahrer solcher Fahrzeuge speziell für einen korrekten Eingriff ausgebildet werden müssen, da sie intuitiv dazu neigen zu früh zu intervenieren.

Um diesem Problem entgegenzuwirken werden vermehrt zusätzliche automatisierte Funktionen eingeführt, um die Bedienung der Systeme zu ermöglichen. Im Bereich des MUM-T werden hierbei Warnungen über Sprachsyntheseprogramme und automatisierte

Planer eingesetzt, um dem Operateur die Möglichkeit zu geben, den Überblick über die Gesamtsituation zu behalten und aufgrund von adäquatem Feedback der automatisierten Teammitglieder Fehlersituationen frühzeitig zu identifizieren.

Wichtig für die Bereitstellung von solchen Unterstützungsfunktionen ist es hierbei auf die jeweiligen Bedürfnisse des Operateurs einzugehen. Beispielsweise werden die Hinweise eines Navigationssystems, dass der Fahrer an der nächsten Ausfahrt abfahren muss, als störend empfunden, wenn dies dem Fahrer bewusst ist und er sich vielleicht sogar schon auf der Abbiegespur befindet. Die Folge ist, dass die entsprechenden Meldungen deaktiviert werden. Andererseits ist derselbe Hinweis sehr willkommen, wenn die Notwendigkeit zum Abbiegen übersehen wurde. Um zu entscheiden, ob eine Sprachausgabe ausgelöst wird, benötigt die Automation daher Wissen über das sogenannte Situationsbewusstsein (Endsley 2001) des Menschen. Neben dieser Anwendung tritt die gleiche Problematik auch im Bereich der Beanspruchung auf. Hier wäre beispielsweise eine zusätzliche automatisierte Anzeige der aktuellen Geschwindigkeitsbegrenzung denkbar, wenn die Beanspruchung des Fahrers aufgrund schlechter Sicht- und Wetterverhältnisse, sehr hoch ist. Allerdings sollten automatisierte Unterstützungsfunktionen nach Möglichkeit nur eingesetzt werden, wenn die Beanspruchungssituation dies erfordert. Hierdurch können sogenannte *Out-of-the-Loop* Effekte vermieden werden, die beispielsweise dazu führen können, dass der Operateur der Automation blind vertraut und somit die entstandenen Fehler nicht abfangen kann. Im vorangegangenen Beispiel handelt es sich hierbei um eine Geschwindigkeitsübertretung, wenn die automatisiert erkannte Begrenzung fehlerhaft ist.

Das Ziel der Vermeidung solcher automationsinduzierten Fehler wirft die Frage auf, wann eine Beanspruchungssituation zusätzliche Unterstützungsfunktionen durch die Automation erfordert. In der Literatur gibt es zu dieser Fragestellung eine Vielzahl von Messmethoden, mit deren Hilfe Indikatoren für Beanspruchung bestimmt werden können. Grundlage dieser sogenannten psychophysiologischen Messungen sind die Reaktionen des menschlichen Körpers auf die einwirkende Belastung. Die Schwierigkeit hierbei ist, dass die Beanspruchung nicht der einzige Faktor ist, der diese Körperreaktionen beeinflusst. Ein Beispiel hierfür stellt die Messung des Pupillendurchmessers dar. Dies ist ein anerkannter Parameter zur Detektion von Beanspruchungsänderungen, wird allerdings zusätzlich von Änderungen der Lichtverhältnisse beeinflusst.

1.1 Forschungsfrage

Ziel dieser Arbeit ist es Messmethoden auf ihre Eignung zur Bestimmung der Nutzerbeanspruchung in Mensch-Maschine-Systemen zu prüfen. Dabei wird versucht

Methoden zu ermitteln mit deren Hilfe die bestimmte Beanspruchung genutzt werden kann, um die Unterstützungsfunktionen an die menschlichen Bedürfnisse anzupassen.

Somit lauten die zu beantwortenden Forschungsfragestellungen:

- Welche Vor- und Nachteile bringen die unterschiedlichen Möglichkeiten der psychophysiologischen Beanspruchungsermittlung?
- Wie können Methoden der psychophysiologischen Beanspruchungsermittlung dazu genutzt werden nutzeradaptive Automation zu ermöglichen?

Zur Beantwortung der Forschungsfrage wurden experimentelle Untersuchungen in einem MUM-T Missionssimulator durchgeführt, in denen jeweils eine Methode zur Beanspruchungsermittlung genauer betrachtet wird. Die Untersuchung ist dabei aufgeteilt in kontextfreie Ermittlungsmethoden, bei denen kein Domänenwissen notwendig ist, und kontextbehaftete Ermittlungsmethoden, deren Erkennungsleistung durch Verständnis über das Aufgabenspektrum erzeugt wird.

1.2 Überblick über die Arbeit

Für die Beantwortung der Forschungsfrage wurden im Rahmen dieser Arbeit mehrere Experimente durchgeführt.

Die Theorie wird in **Kapitel 2 - Mensch-Maschine-Systeme** eingeführt. Dieses Kapitel beschreibt die beiden beteiligten Akteure dieses Kooperationsverhältnisses, die **Maschine** und ihr gegenüber den **Menschen**. Weiterhin werden klassische Konzepte der Mensch-Maschine-Aufgabenteilung erläutert, sowie Probleme, die beim Einsatz von Automation auftreten können. Anschließend wird die **adaptive Automation** als Möglichkeit angeführt, diese statische Zuteilung aufzubrechen. Des Weiteren wird darauf eingegangen welche **Adaptionsmöglichkeiten** einem adaptiven Automationssystem zur Verfügung stehen und welche **Auslöser** für entsprechende Anpassungen genutzt werden können.

Kapitel 3 - Operationalisierung menschlicher Beanspruchung stellt unterschiedliche Ansätze menschliche Beanspruchung zu analysieren vor. Hierzu werden zuerst allgemeine Modelle dargestellt mit denen Aufgabenkontexte beschrieben werden können, um potenzielle Überbeanspruchungssituation zu identifizieren. Zu jedem Modell wird weiterhin untersucht inwiefern sie in adaptiven Automationssystemen einsetzbar sind oder sogar bereits eingesetzt wurden. Abschließend werden die psychophysiologischen Messmethoden als weitere Möglichkeit zur Bestimmung der Beanspruchung beschrieben und dabei in kontextfreie- und kontextbehaftete Ermittlungsmethoden unterschieden.

In **Kapitel 4 - Untersuchung psychophysiologischer Messmethoden** werden die im Rahmen dieser Arbeit durchgeführten Experimente beschrieben. Hierzu wird zuerst der als Testapplikation verwendete **MUM-T Kampffjetsimulator** beschrieben. Daraufhin werden die **Bewertungskriterien** vorgestellt, anhand derer die Forschungsfrage innerhalb der Einzelexperimente genauer untersucht werden soll, bevor die **vier Experimente** im Detail beschrieben werden. Die Experimente werden dabei erneut in kontextfreie- und kontextbehaftete Ermittlungsmethoden aufgeteilt. Am Ende jedes Experiments wird hierbei eine **Bewertung der Messmethode** auf Basis der Bewertungskriterien abgegeben, um die Verwendung der jeweiligen Methode im Rahmen eines adaptiven Automationssystems zu einzuordnen.

Kapitel 5 - Diskussion stellt die Messmethoden gegenüber und entwickelt ein Konzept, wie die beschriebenen Messmethoden zu einem kohärenten adaptiven Automationssystem zusammengefügt werden können, um die jeweiligen Stärken der Systeme zu nutzen.

Am Ende werden in **Kapitel 6 - Zusammenfassung und Ausblick** die gewonnenen Erkenntnisse rekapituliert und darauf aufbauend weitere Untersuchungsgegenstände von adaptiver Automation präsentiert.

2. Mensch-Maschine-Systeme

Das Zusammenwirken eines Menschen mit einer Automation wird nach (Johannsen 2013) als Mensch-Maschine-System bezeichnet. Die Zielsetzung solcher Systeme ist es, den Menschen in die Lage zu versetzen, Aufgaben durchzuführen, die er allein nicht könnte. Ein sehr offensichtliches Beispiel von Mensch-Maschine-Systemen stellt das Steuern von unterschiedlichen Fortbewegungsmitteln wie Automobilen, Flugzeugen oder Schiffen dar. Allerdings werden im Kontext Mensch-Maschine-Systeme neben diesen klassischen Maschinen auch jegliche anderen Formen von technischen Systemen als Maschine bezeichnet. Hierzu zählen insbesondere auch reine Softwaresysteme. Im Folgenden werden zuerst die beiden Teilnehmer **Maschine** und **Mensch** genauer betrachtet, um daraufhin die Schwierigkeiten bei der Gestaltung von Mensch-Maschine-Systemen genauer darstellen zu können. Hierbei wird zum einen diskutiert, wie entschieden werden kann, welche Teile eines solchen Systems automatisiert werden sollten und zum anderen, welche Probleme durch den Einsatz von Automation entstehen können. Abschließend wird **adaptive Automation** als Lösungsansatz zur Vermeidung eben dieser automationsinduzierten Probleme vorgestellt.

2.1 Der Einsatz von Automation

Automatisierte Systeme bieten vielseitige Einsatzmöglichkeiten. Eine Möglichkeit ist das Ermöglichen neuer Fähigkeiten, wie beispielsweise das Fliegen. Andererseits können automatisierte Systeme dazu dienen einen Menschen in seiner Aufgabendurchführung in die Lage zu versetzen die gleiche Aufgabe effizienter durchzuführen. Bei diesem Zusammenwirken bedient der Mensch eine passend entwickelte Maschine, wobei für jeden Anwendungsfall von neuem entschieden werden muss welche Aufgaben oder Teilaufgaben dem Menschen zukommen und welche der Maschine.

Hierzu ist es notwendig den Prozess der Automatisierung nicht starr zu betrachten, sondern als fließendes Kontinuum, von ‚der Mensch führt die Aufgabe komplett selbstständig aus‘, bis hin zu ‚der Mensch ist nicht involviert‘. Um dieses Spektrum genauer zu beschreiben haben (Sheridan und Verplank 1978) eine Taxonomie von Automationsgraden veröffentlicht, die Sheridan et al. in späteren Veröffentlichungen genauer ausdifferenziert haben (Parasuraman et al. 2000). Die am häufigsten verwendete Taxonomie ist die in Abbildung 1 dargestellte zehn Punkte Skala aus (Parasuraman et al. 2000).

- HIGH
10. The computer decides everything, acts autonomously, ignoring the human.
 9. informs the human only if it, the computer, decides to
 8. informs the human only if asked, or
 7. executes automatically, then necessarily informs the human, and
 6. allows the human a restricted time to veto before automatic execution, or
 5. executes that suggestion if the human approves, or
 4. suggests one alternative
 3. narrows the selection down to a few, or
 2. The computer offers a complete set of decision/action alternatives, or
- LOW
1. The computer offers no assistance: human must take all decisions and actions.

Abbildung 1: Levels of Automation Taxonomie aus (Parasuraman et al. 2000)

Demnach ist unter dem niedrigsten Automationsgrad eine Automation zu verstehen, bei der Mensch alles selbst tun muss, wie beispielsweise eine automatisierte Anzeige. In den folgenden Leveln unterstützt die Automation immer stärker bei der Auswahl der richtigen Handlungsalternative, wobei die Entscheidungsgewalt bis zu Level 5 bei dem Menschen liegt. Ab Level 6 beginnt die Automation selbst Entscheidungen zu treffen und der Mensch muss, auf Basis der vorhandenen Informationen, der Entscheidung der Automation aktiv widersprechen. Während die Automation auf Level 10 vollkommen selbstständig agiert und der Mensch nicht mehr involviert ist. Die Autoren beschreiben bei diesen Leveln der Automation, dass es wichtig ist, hohe Automationsgrade nicht als Ziel für die Entwicklung von Automation zu sehen. Stattdessen fordern sie eine Abwägung im Einzelfall, welches Level an Automation für den betrachteten Anwendungsfall angebracht ist. Neben der Taxonomie existieren zusätzlich noch ähnliche Auflistungen für spezialisierte Aufgabenbereiche. Beispielsweise haben (Frohm et al. 2008) eigene *Levels of Automation* für die Fertigung aufgestellt, bei der zwischen manueller und kognitiver Arbeit unterschieden und jeweils eine sieben Punkte Skala aufgestellt wird.

2.2 Der Mensch als überwachende Instanz

Als Gegenpart zur Automation stellt der Mensch als Entscheidungsträger in einem Mensch-Maschine-System eine flexible Komponente dar. Nach (Rasmussen 1983) kann diese notwendige Flexibilität erreicht werden, indem es drei Kategorien von menschlichem Handeln gibt, wie in Abbildung 2 dargestellt. Auf der untersten Ebene „*Skill-Based Behaviour*“ handelt der Mensch unterbewusst auf Basis von erlernten sensomotorischen Mustern und muss dabei leichte ebenfalls unterbewusste Anpassungen aufgrund

wahrgenommener Informationen durchführen. In diese Kategorie fällt beispielsweise das Führen einer Kaffeetasse zum Mund, um daraus zu trinken. Die Arm- und Handbewegungen bei dieser Tätigkeit laufen unterbewusst ab, ohne dabei den Menschen von einer anderen Aufgabe abzulenken, obwohl jedes Mal Anpassungen aufgrund des sich ändernden Gewichts der Tasse passieren müssen, um das Ziel (den Mund) exakt zu erreichen.

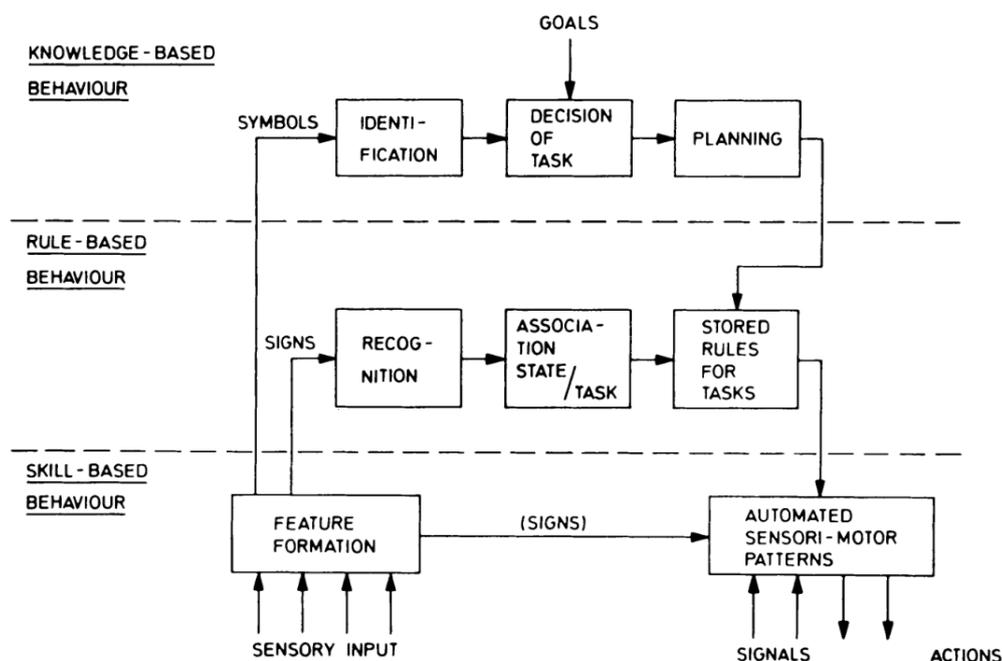


Abbildung 2: Darstellung des Skill-Rule-Knowledge Frameworks aus (Rasmussen 1983)

Die nächsthöhere Ebene stellt das „*Rule-Based Behaviour*“ dar, auf der gespeicherte Regeln abgerufen werden, um ein bestimmtes Ziel zu erreichen. Hierbei wird erst das Problem erkannt und daraus eine Handlung identifiziert, um sie infolgedessen auszuführen. Das Erkennen der zutreffenden Handlung basiert hierbei auf einfachen Regeln, wie beispielsweise das Stoppen vor einer roten Ampel. Auf der höchsten Ebene befindet sich das „*Knowledge-Based Behaviour*“, bei dem der Mensch sein Wissen einsetzen muss, um sich einen Plan zu erstellen. Bei diesem Planungsprozess wird identifiziert, welche Abfolge von Regeln vom aktuellen Zustand zum gewünschten Ziel führen. Somit befindet sich der Entscheidungsprozess im „*Knowledge-Based Behaviour*“, wodurch der Einsatz von adaptiver Automation vorwiegend auf dieser Handlungsebene greifen sollte. Zusätzlich gibt es nach (Parasuraman et al. 2000) vier Informationsverarbeitungsschritte, die im Entscheidungsprozess durchlaufen werden. Diese Schritte sind:

- *Information Acquisition*: die Informationen werden wahrgenommen
- *Information Analysis*: die Informationen werden verstanden
- *Decision Selection*: es wird eine Entscheidung getroffen
- *Action Implementation*: eine Handlung wird auf Basis der Entscheidung ausgeführt

Die adaptive Automation kann hierbei in jedem der einzelnen Schritte eingreifen, wodurch entschieden werden muss, aus welchem Spektrum von Automationsgraden die adaptive Automation in welchem Schritt schöpfen kann. Hierbei ist zu berücksichtigen, dass die Stufen voneinander abhängen. Beispielsweise kann keine sinnvolle Entscheidung getroffen werden, ohne dass der Mensch vorher die zugehörigen Informationen verstanden hat. Aus diesen Abhängigkeiten leiten die Autoren Entwurfsrichtlinien ab, in welcher Stufe wie hoch automatisiert werden kann. Wie in Abbildung 3 dargestellt, kann in den Bereichen *Information Acquisition* und *Information Analysis* ein hoher Automationsgrad gewählt werden, wenn die zugehörige Automation zuverlässig ist. Solch eine Automation beinhaltet automatisierte Aufbereitung von Sensordaten, um beispielsweise eine Zielaufklärung zu vereinfachen indem der Sensor automatisiert ausgerichtet wird (*Information Acquisition*) oder Zielkandidaten salienter dargestellt werden (*Information Analysis*). Im Bereich der *Decision Selection* wird zu hohen Automationsgraden geraten, wenn das Risiko durch eine Fehlentscheidung gering ist. Bei einem hohen Risiko raten die Autoren dazu, nicht über einen automatisierten Vorschlag an den Operateur hinaus zu automatisieren. Für das Ausführen der getroffenen Entscheidung (*Action Implementation*) kann ein hoher Automationsgrad eingesetzt werden. Allerdings wird im Falle einer hochautomatisierten Entscheidung bei einem hohen Risiko durch Fehlentscheidungen zu einem niedrigeren Automationsgrad geraten. Hintergrund hierbei ist die erneute Involvierung des Menschen, um ihn in die Lage zu versetzen, eine Fehlentscheidung im Nachhinein festzustellen und somit rückwirkend zu korrigieren. In diesem „*Error Trapping*“ genannten Ansatz ist die Zielsetzung den Menschen noch einmal zu zwingen, erneut bewusst über eine Entscheidung nachzudenken.

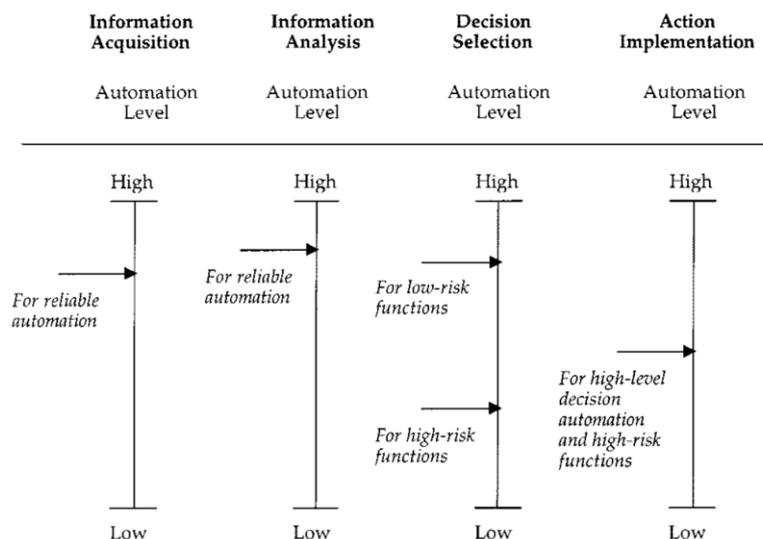


Abbildung 3: Vorschlag für Automationsgrade adaptiver Automation für eine ATC Aufgabe nach (Parasuraman et al. 2000)

2.3 Klassische Mensch-Maschine Aufgabenteilung

Eine der wichtigsten Entscheidungen in der Entwicklung von Mensch-Maschine Systemen ist die Allokation der Aufgaben, d.h. welche Aufgabenbereiche werden von einem automatisierten System durchgeführt, welche Aufgaben bleiben weiterhin beim Menschen und welche Aufgaben werden vom Menschen und der Automation gemeinsam erledigt. Bei der kooperativen Aufgabendurchführung kann nach (Sheridan 1992) hierbei noch unterschieden werden, ob der Mensch in einer manuellen Kontrollsituation ist oder eher in einer überwachenden Kontrollsituation. Bei der manuellen Kontrolle ist der Mensch weiterhin kontinuierlich in die Aufgabendurchführung eingebunden und wird durch eine Automation unterstützt, um die Durchführung zu erleichtern. Ein Beispiel hierfür ist ein Anti-Blockier System im Auto: Der Mensch muss den Bremsvorgang selbstständig initiieren und reguliert ihn kontinuierlich. Die Automation erhöht die dabei mögliche Bremswirkung, indem sie das Blockieren der Reifen verhindert. Bei der überwachenden Kontrolle hingegen, gibt der Mensch eher Zielparametervorgaben an die Automation und überprüft in regelmäßigen Zeitabständen, ob eine Anpassung dieser Zielparameter notwendig ist. Ein Beispiel hierfür stellt der Autopilot eines Flugzeugs dar. Der Pilot gibt die gewünschte Höhe, Geschwindigkeit und Ausrichtung in den Autopiloten ein und muss lediglich in regelmäßigen Abständen überprüfen, ob diese Parameter noch dem aktuellen Ziel entsprechen.

Für die Auswahl, welche Funktionen von einer Automation und welche vom Menschen durchgeführt werden, gibt es nach (Hollnagel und Woods 2005) drei grundlegende Strategien: die Zuweisung nach dem *Left-Over* Prinzip, dem Kompensationsprinzip und dem Komplementärprinzip.

2.3.1 ‚Left-over‘ Zuweisung

Bei der ‚Left-Over‘-Strategie wird jede Funktionalität, die automatisierbar ist, an eine dafür entwickelte Automation übergeben. Automatisierbar bedeutet in diesem Fall, dass es sowohl technisch möglich als auch ökonomisch sinnvoll ist. Dieser Strategie liegt die Annahme zugrunde, dass der Mensch anpassungsfähig ist und somit alle übrigen Aufgaben durchführen kann. Laut (Hollnagel und Woods 2005) trifft diese Annahme allerdings lediglich bei der manuellen Kontrolle zu, in der Regel aber nicht bei der überwachenden Kontrolle, da die hierbei auszuführenden Aufgaben häufig die, in 2.4 beschriebenen, Probleme hervorrufen.

2.3.2 Kompensationsbasierte Zuweisung

Eine Alternative zu der *Left-over* Zuweisung stellt die kompensationsbasierte Zuweisung dar. Bei dieser Strategie wird die Aufgabenallokation anhand eines Vergleichs der Fähigkeiten des Menschen und der Maschine entschieden. Hierzu wird die in Abbildung 4 dargestellte Aufstellung über die jeweiligen Vorteile des Menschen und der Maschine genutzt. Obwohl diese, häufig unter dem Namen *MABA-MABA* (**Men are better at, Machines are better at**) bekannte Liste, bereits 1951, als „*Fitt's List*“, veröffentlicht wurde und seitdem auch häufiger als unzureichend kritisiert wurde (Bye et al. 1999; Clegg et al. 1989; Hoffman et al. 2002), ist sie laut (Winter und Dodou 2014) in ihren Grundzügen bis zur heutigen Zeit immer noch gültig.

Humans appear to surpass present-day machines with respect to the following:

1. Ability to detect small amounts of visual or acoustic energy.
2. Ability to perceive patterns of light or sound.
3. Ability to improvise and use flexible procedures.
4. Ability to store very large amounts of information for long periods and to recall relevant facts at the appropriate time.
5. Ability to reason inductively.
6. Ability to exercise judgment.

Present-day (in 1950s) machines appear to surpass humans with respect to the following:

1. Ability to respond quickly to control signals and to apply great forces smoothly and precisely.
2. Ability to perform repetitive, routine tasks.
3. Ability to store information briefly and then to erase it completely.
4. Ability to reason deductively, including computational ability.
5. Ability to handle highly complex operations, i.e., to do many different things at once.

Note. After Fitts (1951), Hancock and Scallen (1998), and Price (1985).

Abbildung 4: Darstellung der *Fitts List* aus (Inagaki und T)

Dem Menschen werden, bei dieser Strategie, Aufgaben im Bereich der Entscheidungsfindung (6. in der Abbildung) oder des Verbindens abstrakter Konstrukte (4.) zugewiesen. Als Stärken von Maschinen werden das exakte Durchführen immer gleichbleibender Prozesse (2.) oder die gleichzeitige Ausführung unterschiedlicher Aufgaben angegeben (5.).

2.3.3 Komplementäre Zuweisung

Die letzte von Hollnagel beschriebene Strategie sieht überlappende Zuweisungen vor. Somit können die Teilaspekte der auszuführenden Aufgabe sowohl von einer Maschine als auch von einem Menschen durchgeführt werden. Diese Redundanzen sorgen dafür, dass die

Allokation an z.B. den Ausbildungszustand des Menschen oder die vorliegende Situation anpassbar ist. Das Ziel hierbei ist es, langfristig die Leistung der Mensch-Maschine-Kooperation aufrecht zu erhalten, indem bereits zum Zeitpunkt der Entwicklung das Zusammenwirken berücksichtigt wird und somit das Vertrauen des Menschen in die Automation erhöht werden kann.

2.4 Probleme beim Einsatz von Automation

Bei den vorangegangenen Strategien zur Zuweisung von Automation handelte es sich um statische Zuordnungen von Aufgaben an Mensch und Maschine. Hierbei gibt es allerdings eine Vielzahl von menschlichen Faktoren, die zu Problemen durch falsch eingesetzte Automation führen. Die Grundlage dieser Probleme ist die Tatsache, dass die eingesetzte Automation fehlerbehaftet ist, weil bei fehlerloser Automation kein menschlicher Eingriff notwendig oder erwünscht ist, wie beispielsweise bei Anti-Blockier Systemen im Auto. Des Weiteren kommt erschwerend hinzu, dass eine Verbesserung der Automation dazu führen kann, dass die Gesamtleistung des Mensch-Maschine-Systems sinkt. Dieses Phänomen ist bekannt unter dem Namen *Ironies of automation* (Bainbridge 1983). Im Folgenden werden die häufigsten Probleme konventioneller Automation noch einmal genauer diskutiert. Hierbei ist zu beachten, dass sich die einzelnen Probleme nicht gegenseitig ausschließen, sondern sich sogar hervorrufen. Beispielsweise führt *Complacency* häufig zu *Out-of-the-loop* Effekten oder eine schlechte Ressourcenauslastung des Menschen zu *Complacency* und *Skill degradation*.

2.4.1 Skill degradation

Das erste Problem durch konventionelle Automation besteht darin, dass der Mensch durch erhöhte Automation immer seltener selbst die Aufgabe durchführt und dadurch auf lange Sicht die Fähigkeit verliert, die gegebene Aufgabe zu übernehmen, wenn die Automation ausfällt, oder selbst keine adäquate Lösung mehr finden kann. Bereits 1986 wurde dieses Problem im Bereich der psychomotorischen Flugfähigkeiten von Piloten festgestellt (Childs und Spears 1986). Zusätzlich konnten Volz und Dorneich in (Volz et al. 2016) und (Volz und Dorneich 2020) einen erhöhten Verlust von kognitiven Fähigkeiten im Bereich der Flugplanung nachweisen. In den Studien wurden Probanden ohne Vorkenntnisse nach einem eingänglichen Training in drei Gruppen eingeteilt. Eine Gruppe musste die erlernte Flugplanungsaufgabe weiterhin manuell durchführen, wobei die anderen Gruppen zeitweise bzw. permanent durch eine Automation unterstützt wurden. In einem finalen Versuch mussten alle Probanden die Planung erneut manuell durchführen und die Probanden der manuellen Gruppe konnten eine höhere Leistung bei niedrigerer Beanspruchung im Vergleich zu den anderen beiden Gruppen erzielen.

2.4.2 Out-of-the-loop performance

Ein ähnliches Problem, das durch Automatisierung entsteht, ist die *Out-of-the-loop performance*. Komplexe Aufgaben wie beispielsweise Air traffic control erfordern ein mentales Modell der Umgebung, um korrekte Entscheidungen zu treffen. Die Erstellung

eines solchen, oft Situationsbewusstsein genannten (Endsley 1996) mentalen Modells setzt aktive Aufgabendurchführung des menschlichen Operators voraus. Somit kann der Einsatz von Automation das menschliche Situationsbewusstsein negativ beeinflussen, wodurch der Operator als *Out-of-the-loop* bezeichnet wird (Endsley und Kiris 1995). In einem solchen *Out-of-the-loop* Fall benötigt der Mensch länger, um Fehler der Automation festzustellen (Kessel und Wickens 1982) oder bemerkt in Extremfällen nicht einmal, dass die Automation sich deaktiviert hat (U.S. Department of Transportation 1972). Zusätzlich benötigt der Mensch mehr Zeit, um eine Entscheidung zu treffen, da er erst das Situationsbewusstsein wiederaufbauen muss, bevor er handeln kann. Dies stellt ein besonders gravierendes Problem dar, wenn eine Automation ausfällt oder in eine unbekannte Situation kommt und somit der Mensch sehr schnell bereit sein muss, die Kontrolle zu übernehmen (Noy et al. 2018). Dieses Problem zeigt sich derzeit verstärkt im Bereich des autonomen Fahrens von PKW. Hier können die automatisierten Systeme erfolgreich auf einer Autobahn fahren, allerdings muss der Fahrer weiterhin den Verkehr beobachten, um jederzeit die Kontrolle über das Fahrzeug zu übernehmen, wenn die Automation in eine Situation kommt, die sie selbst nicht bewältigen kann.

2.4.3 Trust in Automation

Ein weiteres Problem konventioneller Automation stellt das Vertrauen des Menschen in die Automation dar. Hierbei existieren sowohl Probleme bei einem zu hohen Vertrauen des Menschen in die Automation als auch bei einem zu niedrigen Vertrauen. Zu hohes Vertrauen resultiert hierbei in dem Effekt des *Automation Bias* und zu niedriges Vertrauen äußert sich z.B. in Warnsystemen mit dem sogenannten *Cry Wolf Syndrome*. Probleme dieser Kategorie werden oft auch mit den Begriffen *Overreliance* und *Underreliance* beschrieben (Parasuraman und Riley 1997). Wie in Abbildung 5 dargestellt, verlässt sich der Mensch häufig entweder zu viel (*Misuse*) oder nicht genug (*Disuse*) auf die Automation und somit wird das Potential der Automation nicht vollkommen ausgereizt. Daher ist es notwendig, eine Form von *Calibrated Trust* zu erzeugen, bei dem das Vertrauen des Menschen der Zuverlässigkeit der Automation entspricht. Durch dieses *Calibrated Trust* kann sowohl der *Misuse* als auch der *Disuse* vermieden werden. Hierzu eignet sich laut (Visser et al.) eine Anzeige der Zuverlässigkeit (*Trustworthiness*) der Automation, wie sie in einem Experiment zeigen konnten.

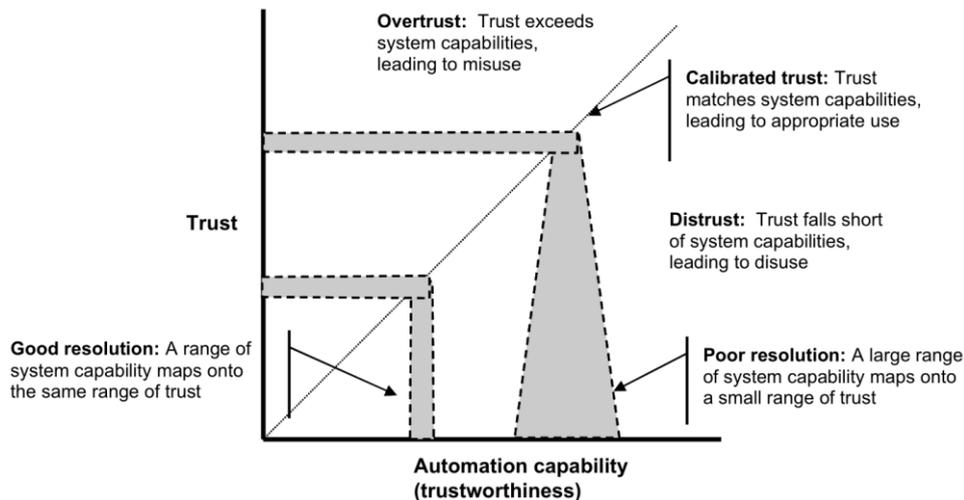


Abbildung 5: Der Zusammenhang von Trust und Automation Trustworthiness aus (Lee und See 2004)

2.4.3.1 Complacency (Automation Bias)

Bei *Complacency* gewöhnt sich der menschliche Operateur an eine gut funktionierende Automation und beginnt deshalb, dieser Automation zuzustimmen oder sich darauf zu verlassen, dass sie ihre Aufgabe korrekt durchführt, selbst wenn diese einen Fehler macht. In (Parasuraman et al. 1993) wurde eine Monitoring Aufgabe in einer Mehrfachaufgabensituation sowohl mit einer konstant niedrigen (87,5%) bzw. hohen (56,25%) Fehlerrate als auch einer variablen Fehlerrate der Automation, in der beide zuvor genannten Fehlerraten auftreten, durchgeführt. Es konnte gezeigt werden, dass bei einer konstanten Fehlerrate die Wahrscheinlichkeit, dass der menschliche Operateur Fehler in der Automation erkennt, deutlich niedriger (82%) ist als bei einer variablen Fehlerrate (33%). Als Vergleich wurde das gleiche Experiment erneut ohne zusätzliche Nebenaufgaben durchgeführt und Automationsfehler wurden mit einer 97%igen Wahrscheinlichkeit erkannt, unabhängig von der Fehlerrate der Automation. In der Literatur wird zusätzlich häufig das Problem „Automation Bias“ beschrieben (vgl. (Cummings) oder (Mosier et al. 1998)), bei dem der Operateur von Vorschlägen der Automation in seinem Entscheidungsprozess beeinflusst wird und dazu tendiert, die vorgeschlagene Entscheidung zu treffen. Allerdings argumentieren (Parasuraman und Manzey 2010), dass beide Begriffe dasselbe Phänomen beschreiben.

2.4.3.2 Cry Wolf Syndrome

Ein gegensätzliches Problem stellt das *Cry Wolf Syndrome* dar. Während bei *Complacency* Fehler der Automation übersehen werden, werden hierbei korrekte Warnungen des Systems ignoriert. Der Name ist angelehnt an eine Fabel über einen Jungen, der vor einem attackierenden Wolf warnt, obwohl kein Wolf anwesend ist. Als der Junge an einem späteren

Zeitpunkt vor einer tatsächlichen Wolfsattacke warnt, glaubt ihm allerdings niemand mehr. Der gleiche Effekt tritt bei automatisierten Warnsystemen ein, die fehlerhafte Warnungen auslösen. Wie in (Sorkin 1988) und (Satchell 2016) berichtet, tendieren Menschen dazu, bei fehlerhaften Warnungen die Warnsysteme zu deaktivieren oder die generierten Warnungen zu ignorieren. Im Gegensatz dazu konnte dieser Effekt in neueren Studien nicht bestätigt werden. (Lim et al. 2019) untersuchten den Effekt von fehlerhaften Tornadowarnungen auf das öffentliche Leben und konnten kein verändertes Verhalten der betroffenen Bürger bei erhöhten Fehlalarmen feststellen. Auch in (Wickens et al. 2009) wurde die Auswirkung von Fehlalarmen, hier in einer Air-Traffic Control Aufgabe, untersucht. Das Ergebnis war erneut, dass eine höhere Rate von falschpositiven Warnungen nicht zu einer reduzierten Leistung der Operatoren geführt hat.

2.4.4 Schlechte Ressourcenauslastung des Menschen

Das letzte Problem bei einer statischen Aufgabenzuweisung sind Schwankungen in der Frequenz der durchzuführenden Aufgaben. Somit kann es zu zeitweise Über- oder Unterforderungen des Menschen kommen, weil sich beispielsweise eine Aufgabe zeitlich stark häuft, wodurch der Operator Unterstützung benötigt oder eine sehr lange Zeit kein Eingriff notwendig ist und der Operator somit die Aufmerksamkeit verliert. Dieser Effekt wird dadurch verstärkt, dass nach (Veltman und Jansen 2005) der Mensch selbst ein adaptives System ist, dass sich an die Begebenheiten der Aufgabensituation anpasst. Ändert sich dabei die Aufgabensituation schlagartig, wie beispielsweise eine dringende Warnung nach einer längeren Ruhephase, muss der Operateur sich an die Situationsänderung anpassen und benötigt zusätzliche Zeit, um die notwendigen Schritte durchzuführen.

2.5 Adaptive Automation

Die Auswahl eines adäquaten Automationsgrads kann, wie bereits in Kapitel 2.3.3 angedeutet, erreicht werden indem für dieselbe Aufgabe mehrere Formen von Automation zur Verfügung gestellt werden und der menschliche Operator selbstständig die passende Automation auswählt. Dieser Ansatz wird als **Adaptierbare Automation** bezeichnet, hat allerdings laut (Prinzel III und Kaber 2006) das Problem, dass Menschen schlecht darin sind zu beurteilen, wieviel automatisierte Unterstützung angebracht ist. Der deshalb weiter verbreitete Ansatz ist die sogenannte **Adaptive Automation**, bei der eine Automation das passende Maß an automatisierter Unterstützung bestimmt. Für den Einsatz von adaptiver Automation gibt es im Wesentlichen zwei Entscheidungsschwerpunkte, die im Folgenden diskutiert werden. Zum einen stellt sich die Frage, welche Anpassungen genutzt werden können, um eine erhöhte Mensch-Maschine-Gesamtleistung zu erzeugen und zum anderen muss entschieden werden, wodurch eine Anpassung ausgelöst wird.

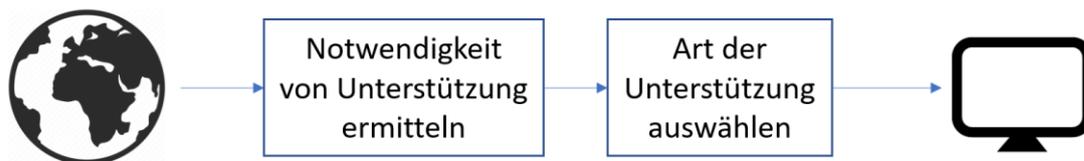


Abbildung 6: Schematische Darstellung von adaptiver Automation

Wie in Abbildung 6 dargestellt werden für die Ermittlung der Notwendigkeit von adaptiven Unterstützungsleistungen Informationen aus der Umwelt des automatisierten Systems verarbeitet. Dieses Verarbeiten kann entweder eine Analyse der Aufgabe beziehungsweise des Arbeitsplatzes oder aber eine Modellierung des menschlichen Operators sein. Sobald ein Unterstützungsbedarf festgestellt wird, muss ein nachfolgendes Modul entscheiden wie die zugehörige Unterstützungsleistung bestmöglich erzielt werden kann. Die Unterstützung wird dabei durch eine Anpassung der Mensch-Maschine Schnittstelle realisiert

2.5.1 Adaptionmöglichkeiten

Nach der Definition von adaptiver Automation soll im Folgenden darauf eingegangen werden, wie eine Adaption stattfinden kann. Diese Adaptionen lassen sich grundsätzlich in zwei Kategorien einteilen. Zum einen kann die Mensch-Maschine-Schnittstelle angepasst werden, um den menschlichen Operator bei seiner Aufgabendurchführung zu unterstützen. Zum anderen können Teilaufgaben zeitweise an eine Automation oder einen anderen Operator übergeben werden. Welche Art der Adaption gewählt werden sollte, hängt

allerdings von dem Problem ab, welches behoben werden soll und wird somit in Kapitel 2.5.2 behandelt.

2.5.1.1 Veränderung der Mensch-Maschine Schnittstelle

Eine Form von Adaptionen ist eine Anpassung der Mensch-Maschine Schnittstelle. Hierbei werden bestimmte, von der Automation für relevant gehaltene, Informationen salienter dargestellt oder weniger relevante Informationen zurückgehalten. Eine weitere Möglichkeit stellt eine Änderung der Darbietungsmodalität dar. Zusätzlich ist eine verbesserte Aufbereitung der vorhandenen Informationen und dadurch die Generierung zusätzlichen Wissens möglich.

Attention Guidance

Bei der Attention Guidance passt die Automation die Darstellung auf der Benutzerschnittstelle an, um relevante Informationen herausstechen zu lassen. In (Horrey und Wickens 2001) konnte die Leistung der Probanden in einer Gefechtsfeldanalyse-Aufgabe durch eine solche Aufmerksamkeitslenkung erhöht werden. In dem Experiment wurden mehrere taktische Elemente auf einer Karte dargestellt. Die Probanden mussten anhand von Parametern wie Gefahr und Zuverlässigkeit der Informationen befreundete Truppen kommandieren, um die Gefahren adäquat zu adressieren. In dieser Aufgabe konnte eine erhöhte Leistung erzeugt werden, wenn die Probanden von einer Automation unterstützt wurden, die wichtige Elemente optisch pulsieren ließ. In einem anderen Anwendungsfall konnten (Schmidt und Rittger 2016) schnellere Reaktionen auf eine Gefahrensituation in einem simulierten Verkehrsszenario erreichen, indem sie die Probanden mittels einer LED auf die Gefahrensituation aufmerksam gemacht haben.

Informationen zurückhalten

Eine weitere Möglichkeit für Adaptionen durch Automation stellt das Zurückhalten irrelevanter Informationen dar. Die Durchführung von wissensbasierten Aufgaben benötigt ein gewisses Maß an fokussierter Aufmerksamkeit. Diese fokussierte Aufmerksamkeit wird häufig als „kognitiver Tunnel“ (siehe Kapitel 2.5.2.5) bezeichnet und wird durch eine Abwesenheit von Ablenkungen begünstigt. Somit können Informationen die den aktuellen Aufgabenkontext nicht beeinflussen, ein erhöhtes Stresslevel auslösen (Mark et al. 2008). Aus diesem Grund ist es laut (McFarlane und Latorella 2002) notwendig eine Form von „Interruption Management“ (Unterbrechungsverwaltung) bei dem Design von Mensch-Maschine Systemen zu berücksichtigen. Im Rahmen dieser Vorstellung konnte in (Dorneich et al. 2012) die Leistung von bodengetriebenen Soldaten erhöht werden. Bei dem Versuch wurden die Soldaten an einen mobilen Elektroenzephalografen angeschlossen, um eine Beanspruchungsbewertung der Soldaten durchzuführen, während sie fünf unterschiedliche

Aufgaben parallel ausführen mussten. Untersucht wurde die Auswirkung eines „*Communications Schedulers*“ der in Situationen hoher Beanspruchung unwichtige Funkmeldungen in eine Textmeldung auf einem Tablet des Soldaten umwandelt. Diese Umwandlung stellt eine Vermeidung von zusätzlichen Ablenkungen dar und konnte eine erhöhte Leistung der Soldaten erzeugen, wenn der „*Communications Scheduler*“ aktiviert war. In einer anderen Anwendung konnten (Theißing und Schulte 2016) automationsinduzierte Fehler verringern, indem das eingesetzte Assistenzsystem „minimalintrusiv“ assistiert hat. Bei diesem Vorgehen nutzt das automatisierte System ein Modell über den menschlichen Entscheidungsprozesses, um zu entscheiden, in welchem Maße eingegriffen werden muss und welche Informationen dabei notwendig sind, anstatt bei jeder erkannten Fehlersituation maximal zu assistieren.

Änderung der Darbietungsmodalität

In den Einsatzgebieten von adaptiver Automation gibt es häufig mehrere unabhängige Informationsquellen, die simultan ausgewertet werden müssen, um ein Situationsbewusstsein aufzubauen. Hierbei kann es zu dem Problem kommen, dass mehrere Informationsquellen in einen Konflikt treten, weil sie beispielsweise beide eine visuelle Informationsaufnahme erfordern (vgl. Kapitel 3.1.3). Aus diesem Grund kann es notwendig sein, die Darbietungsmodalität einer Information anzupassen, um somit die Erstellung des Situationsbewusstseins zu vereinfachen. Hierbei gibt es nach (Schmorrow et al. 2006) die Möglichkeit, die Information in mehreren Modalitäten gleichzeitig darzustellen („*Modality redundancy*“) oder von einer Modalität (z.B. visuell) auf eine andere (z.B. auditiv) zu wechseln („*Modality switching*“). Voraussetzung hierfür ist, dass die Information in mehreren Modalitäten präsentierbar ist. In (Maiwald und Schulte 2016) wurde dieser Ansatz mit einer Erkennung der beanspruchten Modalitäten verbunden, um die Sicherheit von Piloten in einer simulierten Helikoptermission zu erhöhen. In dem beschriebenen Experiment wurde durch die Modalitätserkennung eine unbeanspruchte Modalität ausgewählt, um auf dieser automatisierte Warnungen auszugeben, damit diese besser wahrgenommen werden können.

2.5.1.2 Veränderung des Automationsgrads

Eine weitere Form von Adaptionen stellt das Verändern des Automationsgrads dar. Hierbei werden dem Operateur zusätzliche Automationsfunktionen bereitgestellt, wenn er überlastet ist, oder im Gegenteil in einer Unterbeanspruchungssituation Automationsfunktionen entzogen. Diese zwei Varianten können auch kombiniert werden, indem der standardmäßige Automationsgrad bewusst niedrig gehalten wird und erhöht wird, wenn das automatisierte System einen Bedarf feststellt. Der niedrige initiale Automationsgrad vermeidet hierbei den

Fertigkeitsverlust des Operators und erhöht die Wahrscheinlichkeit eines hohen Situationsbewusstseins. Dennoch können kritische Überlastungssituationen abgefangen werden, indem die zusätzliche Automation in diesen Situationen agiert.

2.5.2 Auslöser für Adaptionen

Neben den Möglichkeiten zur Adaption muss definiert werden, wodurch eine Adaption ausgelöst wird. Hierzu gibt es ein weites Spektrum von Parametern, die meist auf einen Zusammenhang mit Leistungsdekrementen von menschlichen Operatoren zurückzuführen sind.

2.5.2.1 Zeitgesteuerte Adaption

Eine Strategie zur automatisierten Adaption ist eine Änderung des Automationsgrads an festgelegten oder auch zufälligen Zeitpunkten. Diese Strategie wurde in (Kaber und Endsley 2004) eingesetzt, um die Auswirkungen von adaptiver Automation auf einen menschlichen Operator zu evaluieren. Allerdings konnten (Parasuraman et al. 1996) feststellen, dass ihre Versuchsteilnehmer nach einer 10-minütigen Rückkehr zu ihrer Aufgabe eine deutlich höhere Fehlerdetektionsrate aufwiesen. Somit kann auch eine zufällige Adaption des Automationsgrades bereits ein einfaches Mittel zur Verringerung von Leistungseinbußen beim Menschen darstellen.

2.5.2.2 Leistung

Das Ziel von adaptiver Automation ist die Leistung des Gesamtsystems zu erhöhen. Aus diesem Grund ist ein Abfallen der Leistung ein möglicher Auslöser für Adaptionen. Hierbei wird unterschieden zwischen der Leistung in der Hauptaufgabe oder in einer Nebenaufgabe.

Die Leistung in Nebenaufgaben wird häufig verwendet, um die Beanspruchung des Operators festzustellen (Kaber und Endsley 2004; Son und Park 2011; Parasuraman et al. 2009) und somit mögliche Ansatzpunkte für Adaptionen zu identifizieren. Hierbei liegt die Annahme zugrunde, dass eine hohe Leistung in der Nebenaufgabe bedeutet, dass der Operator die Aufgabensituation trotz zusätzlicher Nebenaufgabe gut bewältigen kann und somit auch die Leistung in der Hauptaufgabe als hoch angenommen werden kann. Diese Annahme wird allerdings in (Sauer et al. 2017) insbesondere für Mehraufgabenkontexte in Frage gestellt. Die Autoren argumentieren, dass eine hohe Leistung in der Nebenaufgabe auch bedeuten kann, dass der Operator nicht weiß, wie die Probleme in der Hauptaufgabe zu lösen sind und sich somit auf die Nebenaufgabe konzentriert. Für den Einsatz außerhalb von Experimentalumgebungen besteht darüber hinaus die Schwierigkeit, dass eine

zusätzliche Nebenaufgabe auch eine Zusatzbeanspruchung für den Operateur darstellt und somit die Leistung senken kann. Dieser Effekt wird verstärkt, wenn es sich um eine *künstliche Nebenaufgabe* handelt. Im Gegensatz zu *eingebetteten Nebenaufgaben* passen *künstliche Nebenaufgaben* nicht in den Kontext der Hauptaufgabe, wie beispielsweise eine Kopfrechenaufgabe während Planungsvorgängen. Trotzdem kann eine zusätzliche Nebenaufgabe auch in reellen Einsatzgebieten ein probates Mittel darstellen. Ein Beispiel hierfür, das entfernt an adaptive Automation grenzt, stellt der Totmannschalter für Zugführer dar. Dieser Schalter muss in regelmäßigen Abständen betätigt werden, um zu verifizieren, dass der Zugführer seine Aufgabe weiterhin durchführen kann. Wird der Schalter nicht betätigt, stoppt der Zug automatisch, um einen Unfall zu verhindern.

Auch Leistung in der Hauptaufgabe als Auslöser für adaptive Automation wurde bereits in vielen Veröffentlichungen erfolgreich eingesetzt. Beispielsweise konnte in den zwei aufeinanderfolgenden Publikationen (Calhoun et al. 2012; Calhoun et al. 2011) gezeigt werden, dass die Leistung der Probanden erhöht werden konnte, wenn zusätzliche Automation bereitgestellt wurde, sobald Leistungsschwellwerte unterschritten wurden. Bei der Aufgabe handelte es sich um eine Bildverarbeitung, bei der die Probanden die Anzahl von Rauten in einem Bild ermitteln und über eine Mensch-Maschine Schnittstelle in das System eingeben mussten. Die automatisierte Unterstützung wurde in diesem Szenario durch eine simulierte Klassifikationssoftware, die eine eigene Schätzung als Vorschlag anzeigen konnte, bereitgestellt. Zusätzlich gab es eine weitere Eskalationsstufe, bei der die Schätzung der Software die einzige Auswahlmöglichkeit war und der Proband lediglich die Schätzung bestätigen oder ablehnen konnte. Wie in (Parasuraman et al. 2009) beschrieben, ist es allerdings nicht ausreichend, dass adaptive Automation eine höhere Leistung erzielen kann, als eine Konfiguration ohne Automation. Die adaptive Automation muss zusätzlich eine höhere Leistung als eine Konfiguration mit maximaler Unterstützung erreichen. Dies konnten sie in einem Experiment zeigen, bei dem ihre Probanden unbemannte Fahrzeuge auf einem simulierten Einsatzfeld befehligen mussten. Die Probanden mussten unter anderem Veränderungen in der taktischen Lage („*Change Detection*“) mit einem Tastendruck bestätigen und eine Klassifikation von Bildern durchführen (Ziel oder Nicht-Ziel) die von einem simulierten unbemannten Flugzeug (UAV) gesendet wurden. Für die Klassifikation stand hierbei eine automatisierte Zielerkennung (Automated Target Recognition; ATR) zur Verfügung, die die Bilder bereits vorklassifiziert hat. In einem Vergleich von den folgenden drei Automationsstrategien:

- **Manuell**, die ATR steht nicht zur Verfügung
- **Statisch**, die ATR wird nach der Hälfte der Versuchsdauer aktiviert
- **Adaptiv**, die ATR wird aktiviert, wenn die Leistung in der *Change Detection* Aufgabe niedrig ist

konnte die adaptive Strategie eine höhere Leistung in sowohl der assistierten Aufgabe als auch allen nicht assistierten Aufgaben hervorrufen als die beiden anderen Strategien. Somit konnte gezeigt werden, dass adaptive Automation auf Basis von Leistung in der Hauptaufgabe zu einer Leistungserhöhung des Mensch-Maschine Gesamtsystems gegenüber sowohl manueller als auch der vollständig assistierten Aufgabendurchführung führen kann.

Allerdings ist der Einsatz von Leistung in der Hauptaufgabe in realen Aufgabengebieten ebenfalls problematisch. Oftmals kann die Leistung nicht zuverlässig gemessen werden, vor allem, weil eine solche Leistungsmessung voraussetzt, dass die Aufgabe vollends von einer Automation durchgeführt werden kann, wodurch der Einsatz eines Mensch-Maschine Systems nicht notwendig wäre. Aus diesem Grund werden häufig nur Leistungsindikatoren herangezogen, wie beispielsweise die Anzahl an noch nicht bearbeiteten. Des Weiteren argumentiert (Scerbo 2001), dass adaptive Automation auf Basis von Leistung Leistungsdekremente voraussetzt und somit „zu spät“ assistiert, nämlich erst nach dem Auftreten potentiell kritischer Fehler.

2.5.2.3 Situationsbasiert

Der naheliegendste Auslöser für Adaptionen ist eine Interpretation der in der Mensch-Maschine Schnittstelle visualisierten Daten. Die Notwendigkeit von zusätzlicher Automation beruht zumeist auf einer Veränderung der Aufgabensituation, die eine hohe Belastung auf den Operateur auslöst. Dieses Wissen kann genutzt werden, um Situationen hoher Belastung vorab zu modellieren, damit ein automatisiertes System diese während des Einsatzes des Mensch-Maschine Systems erkennen kann und somit die Automation entsprechend anpassen kann. Laut (Lotz et al. 2020) ist es außerdem möglich, diese kritischen Situationen mit Hilfe von maschinellen Lernverfahren auf Basis von realen Daten abzubilden, um diese dann wieder zu identifizieren. Sie versprechen sich von einem solchen Ansatz ein erhöhtes Vertrauen und somit eine Minderung des ‚Disuse‘ von Automation (vgl. Kapitel 2.4.3). Dieses erhöhte Vertrauen auf Basis von Adaption in kritischen Situationen konnten (Muslim und Itoh 2019) in einem Verkehrssimulationsexperiment nachweisen. In ihrem Experiment wurden Autofahrer bei Spurwechselmanövern unterstützt, indem entweder die Trägheit des Lenkrads erhöht wurde oder das automatisierte System die Kontrolle über die laterale Geschwindigkeit über das simulierte Fahrzeug übernommen hat, wenn eine von vier vordefinierten Gefahrensituationen erkannt wurde. Diese Gefahrensituationen bestanden aus Fahrzeugen, die sich auf der Zielspur in den Bereichen:

- *Front-Proximity Zone (FPZ)*: Das andere Fahrzeug ist im vorderen Bereich des eigenen Fahrzeugs und kann somit gesehen werden.

- *Blind-Spot* (BS): Das andere Fahrzeug befindet sich im toten Winkel des eigenen Fahrzeugs und kann nur über einen Schulterblick gesehen werden.
- *Rear-Proximity Zone* (RPZ): Das andere Fahrzeug ist im hinteren Bereich des eigenen Fahrzeugs und somit im Außenspiegel sichtbar.
- *Fast Approach Zone* (FAZ): Das andere Fahrzeug ist hinter dem eigenen Fahrzeug, aber hat eine erhöhte Geschwindigkeit. Es ist im Außenspiegel sichtbar.

befinden, die in Abbildung 7 dargestellt sind.

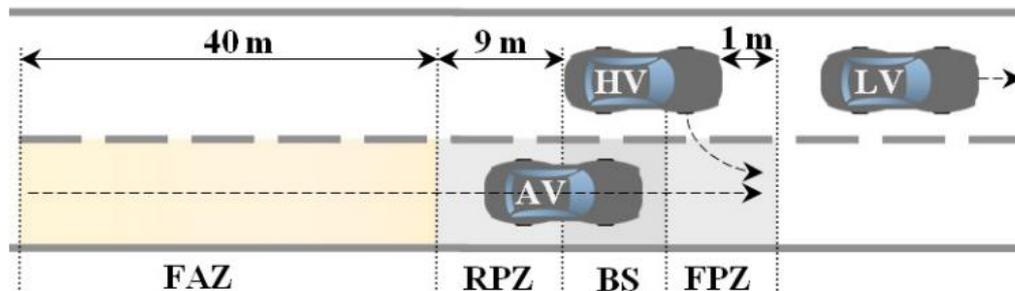


Abbildung 7: Gefahrenzonen in der Fahrzeugsimulation von (Muslim und Itoh 2019)

Wird ein Fahrzeug während eines Spurwechsels in einem dieser Bereiche erkannt, wird die automatisierte Unterstützungsfunktion aktiviert. Dies wurde verglichen mit einer adaptiven Automation, bei der beide genannten Unterstützungsfunktionen aktiviert werden konnten und die Automation je nach Situation unterschiedliche Maßnahmen gewählt hat. Hierbei wurden automatisierte Funktionen lediglich genutzt, wenn das erkannte Risiko für den Fahrer schwer ersichtlich war. Hierdurch konnte eine höhere Akzeptanz und ein erhöhtes Vertrauen in die Unterstützungsfunktionen erzielt werden als bei den nicht-adaptiven Automationsstrategien.

Das beschriebene Experiment zeigt die Schwierigkeiten des situationsbasierten Auslösens von Automation. Zum einen müssen potenzielle Gefahrenquellen im Vorfeld modelliert werden und zum anderen müssen die Eingriffe auf tatsächliche Gefahrensituationen begrenzt werden, um das Vertrauen in die Automation aufrechtzuerhalten. Erschwerend kommt hinzu, dass die Bewertung der Situation vom individuellen Menschen abhängt. In Situationen, die beispielsweise ein erfahrener Mensch noch selbstständig bewältigen möchte, wünscht sich ein Novize bereits die größtmögliche Unterstützung. Diese Unterschiede kann eine situationsbasierte Adaption nicht berücksichtigen, da die auf den Menschen wirkende Belastung analysiert wird und nicht die daraus resultierende Beanspruchung.

2.5.2.4 Kognitionsergonomische Konstrukte

Eine Alternative zu den bisherigen Auslösern, die vollständig auf objektiv messbare Größen zurückführbar sind, ist eine Anpassung an den mentalen Zustand des menschlichen

Operateurs. Der Vorteil dieser Methode ist, dass individuelle Unterschiede in der Bearbeitung der notwendigen Aufgaben genutzt werden können, um Auslöser zu ermitteln, die besser auf den Operateur zugeschnitten sind. Ein intuitives Beispiel für eine solche direktere Anpassung ist ein Navigationssystem im Auto, welches erkennt, dass der Fahrer bereits dabei ist von einer Autobahn abzufahren und deshalb die Meldung unterdrückt, die dem Operateur mitteilt, dass er abfahren muss.

Der mentale Zustand des Operateurs ist allerdings nicht direkt messbar, sondern kann ausschließlich über die indirekte Messungen inferiert werden. Auf Basis der Messungen kann der mentale Zustand auf unterschiedliche Gesichtspunkte hin untersucht werden. Für diese Untersuchung werden Konstrukte aus der Kognitionspsychologie genutzt, mithilfe derer die Art und Weise, wie die Aufgabendurchführung stattfindet, genauer beschrieben werden kann. Das vorangegangene Beispiel bildet eine automatisierte Inferenz der Aufmerksamkeit ab, aber auch das vorhandene Wissen des Operateurs oder die Beanspruchung bei der Aufgabendurchführung können für adaptive Automation genutzt werden. Die angesprochenen Konstrukte werden im Folgenden genauer diskutiert und dargestellt, wie diese für adaptive Automation genutzt werden können.

2.5.2.5 Aufmerksamkeit

Die Aufmerksamkeit beschreibt die Aufgabe oder Information, mit der der Mensch sich zu einem Zeitpunkt beschäftigt. Laut (Duchowski 2003) ist dieser Prozess bedingt durch die Limitationen der Informationsverarbeitung des menschlichen Gehirns. Er beschreibt weiterhin:

„The brain processes sensory input by concentrating on specific components of the entire sensory realm so that interesting sights, sounds, smells, etc., may be examined with greater attention to detail than peripheral stimuli.“ (Duchowski 2003)

Somit stellt Aufmerksamkeit eine Vorverarbeitung der vorhandenen Informationen dar, die das Ziel hat, diese auf das Wesentliche zu reduzieren. Auf den ersten Blick erlaubt diese Reduktion somit eine erhöhte Aufnahme relevanter Informationen durch das Unterdrücken irrelevanter Informationen. Allerdings bestehen die Einsatzgebiete adaptiver Automation aus mehreren simultanen Aufgaben, wodurch sich regelmäßig wandelt, welche Informationen relevant und welche irrelevant sind. Für diese Art von Einsatzgebieten ist eine Form von *geteilter Aufmerksamkeit* („divided attention“) notwendig, bei der unabhängige Informationen gleichzeitig aufgenommen werden. Wie (Broadbent 1957) in einem Experiment, in dem Zahlen rezitiert werden mussten, die gleichzeitig auf jeweils nur einem Ohr abgespielt wurden, gezeigt hat, ist dies zwar möglich, allerdings nur unter starken Leistungseinbußen.

Statt der geteilten Aufmerksamkeit tendieren Menschen daher dazu, Informationen sequenziell aufzunehmen. Diese *selektive Aufmerksamkeit* („selective attention“) existiert nicht nur für die Informationsaufnahme, sondern auch für das Durchführen von Aufgaben (Kerstholt und Passenier 2000; Boehne und Paese 2000). Bei der *selektiven Aufmerksamkeit* kann die Informationsverarbeitung vollkommen auf die aktuelle Aufgabe ausgerichtet werden, um das bereits beschriebene Unterdrücken irrelevanter Informationen zu erzielen. In Situationen besonders hoher Konzentration können hierbei sogar sehr offensichtliche Dinge übersehen werden wie im berühmten Gorilla Experiment von (Simons und Chabris 1999). Bei dem Experiment haben die Probanden ein Video angesehen, bei dem eine Gruppe von Menschen sich einen Ball zugeworfen und sich dabei bewegt hat, sodass regelmäßig die Positionen der Menschen variiert wurde. Hierbei wurden die Probanden gebeten zu zählen, wie häufig der Ball geworfen wurde. Die Aufmerksamkeit auf den Ball hat hierbei bei der Hälfte der Probanden dazu geführt, dass sie nicht bemerkt haben, dass währenddessen eine Person in einem Gorillakostüm durch die Menschengruppe gegangen ist. In der Literatur wird dieses Phänomen als *Unaufmerksamkeitsblindheit* („inattentional blindness“) bezeichnet. Umgekehrt wird das starke Fokussieren auf die gegebene Aufgabe, sodass *Unaufmerksamkeitsblindheit* auftreten kann, als *Aufmerksamkeitstunnel* („attentional tunneling“) oder *kognitiver Tunnel* („cognitive tunnel vision“) bezeichnet (Wickens; Moray). Nach (Staal 2004) kann dieser *Aufmerksamkeitstunnel* zu einer Leistungserhöhung führen, wenn die zugehörige Aufgabe keine periphere Informationsverarbeitung benutzt, allerdings führt es laut (Neerinx 2003) auch dazu, dass Menschen dazu neigen, sich auf eine Aufgabe zu fokussieren, selbst wenn eine andere Aufgabe wichtiger ist. Dieses Zögern, die Aufgabe zu wechseln, wird weiterhin verstärkt, indem das Wechseln von einem Aufgabenkontext in einen anderen zusätzliche „*task-switching costs*“ mit sich bringt. Diese zusätzliche mentale Beanspruchung kann ebenfalls auftreten, wenn der Mensch von der Bearbeitung der aktuellen Aufgabe in einen Hilfekontext wechselt, weil eine automatisierte Unterstützungsfunktion ausgelöst wurde (Gopher et al. 2000).

Für die adaptive Automation kann eine Analyse der Aufmerksamkeit des Operators hilfreich sein, weil hierdurch innerhalb des automatisierten Systems verstanden werden kann, ob der Operator kritische Systemzustände erkannt hat und die Behebung der dadurch auftretenden Probleme bearbeitet. Des Weiteren sollte beim Lenken der Aufmerksamkeit des Operators berücksichtigt werden, dass dies einen zusätzlichen Kontextwechsel darstellt, wodurch die Leistung gesenkt werden kann.

2.5.2.6 Situationsbewusstsein

Das Situationsbewusstsein stellt ein weiteres Modell im Bereich der menschlichen Informationsverarbeitung dar. Hierbei handelt es sich um das Verständnis eines Operators

über die gegebene Situation. (Endsley 1988a) beschreibt das Situationsbewusstsein („*Situation Awareness*“) als einen kritischen Input für die menschliche Entscheidungsfindung, die wiederum die Grundlage für alle ausgeführten Aktionen darstellt. Demnach fällen gut ausgebildete Operateure falsche Entscheidungen, wenn sie kein gutes Abbild der Situation besitzen. Abbildung 8 stellt den Zusammenhang zwischen den relevanten Parametern des menschlichen Entscheidungsprozesses dar, wie er im Situationsbewusstseinsmodell angenommen wird. Dabei wird eine Entscheidung auf Basis von Zielen und dem Situationsbewusstsein getroffen, wobei zusätzlich die zugrundeliegende Beanspruchung und individuelle Faktoren wie der Ausbildungszustand des Menschen in die Entscheidungsfindung einfließen. Die getroffene Entscheidung wird daraufhin ausgeführt, wobei erneut Beanspruchung und individuelle Faktoren beeinflussen, wie akkurat die getroffene Entscheidung ausgeführt werden kann. Die durchgeführte Aktion wiederum erzeugt einen Einfluss auf die Umwelt, wodurch das Situationsbewusstsein aktualisiert werden muss und der Prozess erneut startet.

Endsley teilt das Situationsbewusstsein in drei aufeinander aufbauende Stufen ein: „Wahrnehmung“, „Verständnis“ und „Projektion“.

Die erste Stufe der *Wahrnehmung*, oder auch Level-1 Situationsbewusstsein, beschreibt das Wissen über die sichtbaren Elemente in der aktuellen Aufgabendurchführung. Hierzu zählen das Bewusstsein über die Existenz von relevanten Objekten sowie deren Position.

In der Stufe *Verständnis* wird die Bedeutung der wahrgenommenen Objekte bewertet. Level-2 Situationsbewusstsein bedeutet somit beispielsweise, dass ein Operateur eine aktive Warnleuchte wahrgenommen hat und zusätzlich versteht, ob diese Warnung eine Gefahr in der aktuellen Situation darstellt.

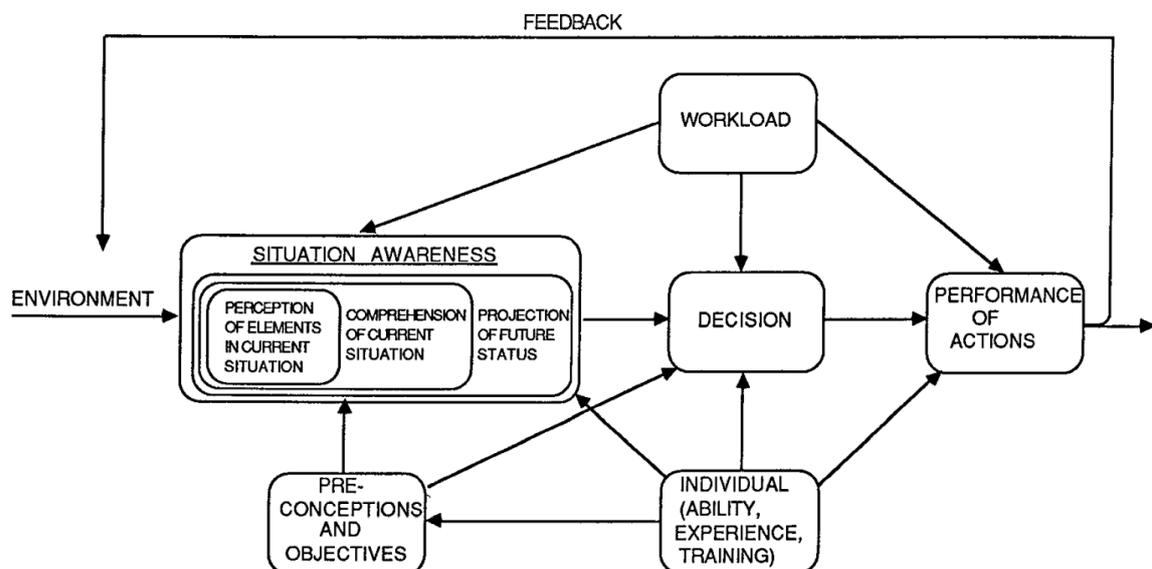


Abbildung 8: Entscheidungsmodell des Menschen aus (Endsley 1988a)

Die letzte Stufe beschreibt die Fähigkeit, die wahrgenommenen Informationen in die Zukunft zu projizieren, um zukünftige Probleme zu antizipieren. Hierbei müssen häufig eine Vielzahl von Informationen mit Domänenwissen verknüpft werden, um korrekte Vorhersagen treffen zu können. Somit beschreibt Level-3 Situationsbewusstsein nicht ausschließlich die Fähigkeit, eine gegebene Situation zu analysieren, sondern auch eine Verbindung mit erlerntem Wissen. Aus diesem Grund kann Level-3 Situationsbewusstsein für gewöhnlich ausschließlich durch einen gesammelten Erfahrungsschatz erreicht werden.

Die Bewertung von Situationsbewusstsein wird bis dato vorwiegend über Fragebogenmethoden wie die Situation Awareness Global Assessment Technique (SAGAT) (Endsley 1988b) oder Situational Awareness Rating Technique (SART) (Taylor 2017) durchgeführt. Dabei wird die Simulation an zufälligen Zeitpunkten angehalten, um die Fragebögen auszufüllen. Alternativ kann eine Echtzeitfeststellung von Situationsbewusstsein durch Befragungen während der Aufgabendurchführung erreicht werden, allerdings erzeugt diese Methode zusätzliche Beanspruchung (Jones und Endsley 2004).

Im Bereich nicht intrusiver Verfahren zur Abschätzung von Situationsbewusstsein wird Blickbewegungsmessung eingesetzt. So haben (Hasanzadeh et al. 2016) ein mobiles Blickbewegungsmesssystem genutzt, um das Situationsbewusstsein von Arbeitern auf einer Baustelle zu ermitteln und das Ergebnis mit Fragebogenmethoden zu vergleichen. Sie kamen zu dem Resultat, dass die Blickbewegungen eine Korrelation zu den Fragebögen aufweisen und die Messung geeignet ist, um den Bedarf von Sicherheitsschulungen für die Bauarbeiter festzustellen.

Der Einsatz von Echtzeitmessungen des Situationsbewusstseins in einem adaptiven Automationssystem wird außerdem parallel zu diesem Promotionsvorhaben derzeit von Simon Schwerd in seinem Promotionsvorhaben untersucht. Wie in (Schwerd und Schulte 2020) beschrieben, bildet er dabei auf Basis von Blickbewegungsdaten und manuellen Interaktionen eine Abschätzung des Level-1 Situationsbewusstseins. Hierbei werden die fixierten Informationen als wahrgenommen bewertet und daraufhin ein Modell des wahrgenommenen Wissens gebildet. Mithilfe dieses Modells wird daraufhin ermittelt, welche Informationen von dem letzten wahrgenommenen Wert abweichen und somit entsteht ein Modell, mit dem man die Aktualität des vorhandenen Wissens betrachten kann.

2.5.2.7 Beanspruchung

Die Aufgaben, die ein Operateur in der Ausführung seines Auftrags durchführen muss, stellen eine *Belastung* dar. Das Bearbeiten dieser Aufgaben und die damit verbundenen Reaktionen im menschlichen Körper werden als *Beanspruchung* bezeichnet. Während die genauen Beschreibungen hierbei auseinander gehen, beinhalten sie alle den Verbrauch

mentaler Ressourcen. Physische Ressourcen können ebenfalls verbraucht werden, die Analyse dieser Ressourcen liegt allerdings außerhalb des Themengebiets dieser Arbeit. Die Stärke der Beanspruchung hängt von einer Vielzahl von Faktoren ab. Einen dieser Faktoren stellt die zugrundeliegende Arbeitslast dar. Zu dieser Arbeitslast gehört die zu bewältigende Aufgabe, aber auch Umweltfaktoren wie beispielsweise Lärm können die Arbeitslast erhöhen. Hierbei gilt grundsätzlich, dass eine höhere Belastung auch eine höhere Beanspruchung bedeutet. Weiterhin hängt die Beanspruchung aber auch von individuellen menschlichen Faktoren wie dem Ausbildungsstand und oder den Fertigkeiten des Menschen ab. Weitere Faktoren stellen Ermüdungszustand oder Emotionslage dar. Zusätzlich ist zu bedenken, dass die Beanspruchung in mehrere Teilaspekte aufgeteilt werden kann. Die Beanspruchung wird häufig in mehrere unterschiedliche mentale Ressourcen unterteilt, wie die visuelle Aufnahmeressource oder die kognitive Ressource. Je nach Beschreibung von Beanspruchung kann demnach entweder eine bestimmte dieser Ressourcen oder aber die Gesamtheit aller Ressourcen gemeinsam gemeint sein.

Während die *Belastung* eine quantitativ messbare Größe darstellt, sind die anderen beschriebenen Faktoren lediglich qualitativ erfassbar. Aus diesem Grund wird *Beanspruchung* als ein Konstrukt bezeichnet, welches nicht direkt messbar ist. Um dieses Konstrukt dennoch nutzbar zu machen können Modelle genutzt werden, die die Beanspruchung mit messbaren Größen verbindet und sie dadurch *operationalisierbar* macht. Eine Möglichkeit dieser Operationalisierung stellen psychophysiologische Messmethoden dar, bei denen die Reaktionen des menschlichen Körpers auf die *Belastung* gemessen werden, um somit die verbundene Beanspruchung zu schätzen. Diese Messungen werden in Kapitel 3.2.1 genauer diskutiert. Eine weitere Möglichkeit der Operationalisierung von Beanspruchung stellen Beanspruchungsmodelle dar, bei denen Erkenntnisse aus Experimenten genutzt werden, um die relevanten Beanspruchungsparameter zu bestimmen und daraus eine Beanspruchungsschätzung zu generieren. Diese Modelle können daraufhin in einem automatisierten System implementiert werden, sodass es darauf aufbauend in Echtzeit adaptive Automationsentscheidungen treffen kann. Ein großer Mehrwert dieser Beanspruchungsmodellierung ist die Möglichkeit, kritische Aufgabensituationen identifizieren zu können, bevor sie auftreten, indem die aktuelle Situation in die Zukunft projiziert werden kann. Im Folgenden wird erst der Zusammenhang zwischen der Beanspruchung und der davon abhängigen Leistung des Menschen diskutiert und daraufhin werden in Kapitel 3 etablierte Beanspruchungsmodelle zusammen mit bereits erfolgreichen Einsätzen im Bereich der adaptiven Automation vorgestellt.

2.5.2.8 Zusammenhang zwischen Beanspruchung und Leistung

Wie bereits diskutiert ist das Ziel adaptiver Automation, die Leistungsfähigkeit des Menschen zu erhöhen. Aus diesem Grund ist es wichtig zu verstehen, wie sich die Leistung auf Basis der Beanspruchung verändert. Die Grundlage dieser Untersuchungen lässt sich hierbei auf das „*Yerkes-Dodson Law*“ zurückführen (Yerkes und Dodson 1908). Nach diesem Gesetz ist der Zusammenhang zwischen Beanspruchung und Leistung eine *invertierte U-Kurve*, bei der sowohl niedrige als auch hohe Beanspruchungsniveaus zu einer schlechteren Leistung führen als ein mittleres Beanspruchungsniveau. Wichtig ist es hierbei zu erwähnen, dass das ursprüngliche Experiment von Yerkes und Dodson weder etwas mit Beanspruchung noch mit Leistung oder gar mit Menschen zu tun hatte. Stattdessen wurde die Lernrate von Ratten bei einem variablen Level von Erregung („*Arousal*“) untersucht. Zu diesem Zweck mussten die Ratten eine von zwei möglichen Boxen betreten. Wurde die dunklere der zwei Boxen ausgewählt, bekam die Ratte einen Stromstoß versetzt. Hierbei wurde die Stärke der Stromstöße variiert und eine mittlere Stromstoßstärke führte schneller dazu, dass die Ratten drei Tage in Folge zehn Mal hintereinander die richtige Box auswählten, als sowohl eine hohe wie auch eine niedrige Stromstoßstärke. Trotz dieser Diskrepanz zwischen dem Ursprungsexperiment und der verbundenen Aussage wird die *invertierte U-Kurve* dem Zusammenhang zwischen Beanspruchung und Leistung zugrunde gelegt. Allerdings wurde der besagte Zusammenhang um die *Anstrengung* („*Effort*“) des Menschen erweitert. Durch *Anstrengung* kann der Mensch der Aufgabe zusätzliche Aufmerksamkeitsressourcen widmen und dadurch eine Abnahme der Leistung in Situationen hoher Beanspruchung verzögern (Matthews und Davies 2001). Allerdings führt diese Anstrengung zu einer schnelleren Ermüdung und einer längeren Zeit, in der sich der Mensch erholen muss (Veltman und Jansen 2005). Beanspruchung über das durch *Anstrengung* erreichbare Limit hinaus wird als Überlastung bezeichnet, bei der der Mensch dazu tendiert, die bereits aufgewendete *Anstrengung* aufzugeben. Des Weiteren sprechen (Grier et al. 2008) in Multitasking Aufgabensituationen von einer *red-line of workload*, also einer roten Linie der Beanspruchung. Hierbei handelt es sich um ein hohes Beanspruchungslevel, welches bereits vor einer Überlastungssituation zu einem Einbruch der Leistung des Menschen führt. Abbildung 9 stellt die beschriebenen Zusammenhänge noch einmal bildlich dar.

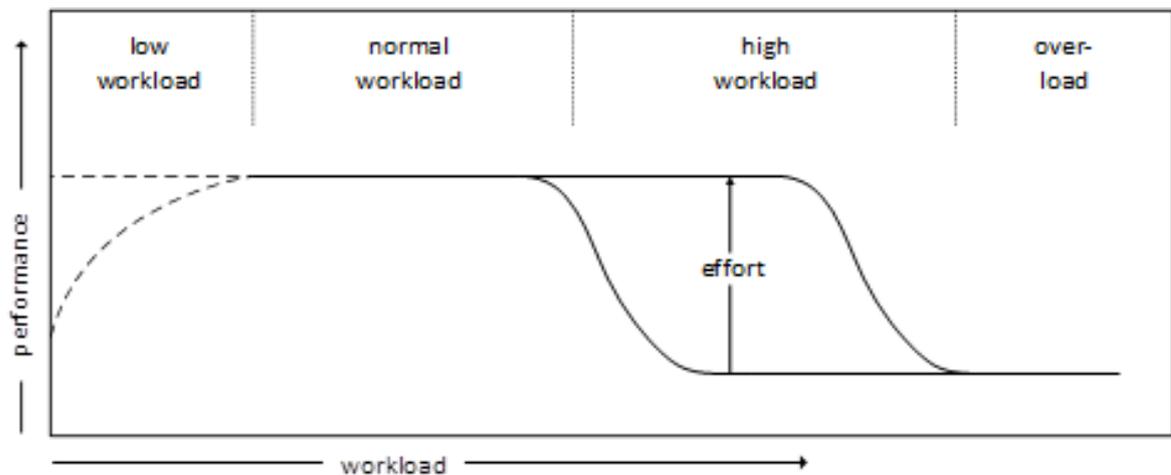


Abbildung 9: Visualisierung des Zusammenhangs zwischen Beanspruchung und Leistung

2.6 Konsolidierung

Der Einsatz von adaptiver Automation kann auf unterschiedliche Arten und Weisen durchgeführt werden. Die Adaptionen können durch eine Anpassung der Darstellung der Informationen oder durch eine Vereinfachung der Aufgabe umgesetzt werden. Zusätzlich kann auch das aktive Zurückhalten von Informationen eine adäquate Adaptionstrategie darstellen, wenn die Darstellung der Information eine Überforderungssituation für den Operateur bedeutet. Die Ermittlung der Adaptionauslöser ist ebenfalls durch eine Vielzahl von unterschiedlichen Optionen möglich. Die Nutzung der menschlichen Leistung oder einer Analyse der Situation liefert ein sehr direktes Maß für die Notwendigkeit von zusätzlicher Automation, bildet gleichzeitig allerdings nur bedingt den tatsächlichen Bedarf ab, weil dieser von Mensch zu Mensch unterschiedlich ist. Daher ist es besser die individuellen Bedürfnisse des Operateurs festzustellen und entsprechend zu unterstützen. Leider sind diese Bedürfnisse nicht direkt messbar und müssen deshalb über kognitionsergonomische Konstrukte indirekt ermittelt werden. Für diese indirekte Ermittlung werden Zusammenhänge zwischen den Konstrukten und messbaren Größen hergestellt, wodurch die Notwendigkeit an automatisierter Unterstützung in Echtzeit ermittelt werden kann. Während jedes der beschriebenen Konstrukte genutzt werden kann, um eine solche menschadaptive Automation bereitzustellen, fokussiert sich diese Arbeit auf das Konstrukt „Beanspruchung“, weshalb sich das folgende Kapitel mit den unterschiedlichen Arten zur Operationalisierung der menschlichen Beanspruchung beschäftigt.

3. Operationalisierung menschlicher Beanspruchung

Bei der Ermittlung menschlicher mentaler Beanspruchung besteht die Schwierigkeit, dass diese keine messbare Größe darstellt. Unterschiedliche Stufen von Aufgabenlast führen je nach Erfahrung und Trainingsstatus beziehungsweise sogar je nach Gemütszustand des menschlichen Operateurs zu unterschiedlichen Leveln an Beanspruchung, wodurch diese eine sehr individuelle Größe darstellt. Dennoch können außenstehende Menschen durch eine Beobachtung durchaus feststellen, wann ein Mensch überlastet ist oder zusätzliche Unterstützung benötigt. Aus diesem Grund beschäftigt sich die Kognitionspsychologie mit der Operationalisierung der Beanspruchung, also der Inferenz des Beanspruchungszustands aus indirekten Messungen. Eine Variante für die Operationalisierung stellt die Modellierung der Beanspruchung durch z.B. eine Beschreibung der durchzuführenden Aufgaben oder aber durch eine strukturierte retroaktive Befragung der Probanden dar. Als Alternative dazu kann die Beanspruchung auch mit der Reaktion des menschlichen Körpers auf die Belastung abgeschätzt werden. Dabei werden psychophysiologische Messmethoden eingesetzt, um die körperlichen Reaktionen aufzunehmen, die zuvor bereits mit einer Veränderung der Beanspruchung in einen Zusammenhang gestellt werden konnten. In der Literatur beziehen sich die Messungen dabei grundsätzlich auf das Auge, das Herz oder das Gehirn.

Dieses Kapitel stellt beide Ansätze noch einmal im Detail vor und geht dabei erneut darauf ein, inwiefern sich die jeweiligen Methoden in eine closed-loop Anwendung einbinden lassen. Für die Modellansätze besteht dabei das Problem, dass abstrakte Beschreibungen der Tätigkeiten nur schwierig in einem automatisierten System realisiert werden können, während für die psychophysiologischen Messmethoden häufig unklar ist, ob die Veränderung der Messgröße tatsächlich auf eine erhöhte Beanspruchung zurückzuführen ist.

3.1 Modellierung von Beanspruchung

Ein Kernaspekt für ein erfolgreiches Zusammenwirken zwischen Mensch und Maschine ist die Modellierung der menschlichen Beanspruchung, um dadurch die Automation so zu gestalten, dass sie auf die Stärken und Schwächen des Operateurs eingeht. Zu dieser Anpassung gehören Designentscheidungen in der Mensch-Maschine-Aufgabenteilung sowie der Einsatz zusätzlicher Messmethoden, um in dem automatisierten System ein Verständnis für kritische Situationen zu erzeugen. Im Rahmen dieser Arbeit werden unter solchen kritischen Situationen vor allem Situationen betrachtet, in denen der Mensch überlastet ist. In der Literatur existieren unterschiedliche Modelle zur Berücksichtigung von Beanspruchung, die es ermöglichen sollen, potentielle menschliche Fehler vorhersagen zu können. Die häufig in der Literatur diskutierten Modelle bestehen aus Fragebogenmethoden wie dem NASA-TLX, Modellen zur Beschreibung von menschlichen Tätigkeitsprofilen

(z.B. Single Resource Theory oder Cognitive Task-Load Modell) und Methoden zur Ermittlung der Beanspruchung in Echtzeit. Dieses Kapitel stellt diese Modelle, sowie deren Einsatz in adaptiven Systemen werden im Folgenden vorgestellt.

3.1.1 Kognitive Architekturen

Ein besseres Verständnis für die menschliche Informationsverarbeitung kann einem automatisierten System helfen, potenzielle Fehlerfälle zu identifizieren. Zu diesem Zweck können die sogenannten kognitiven Architekturen eingesetzt werden. Diese Stellen Frameworks zur Verfügung, mit denen Teile des menschlichen Kognitionsprozesses nachgebildet werden können. Die beiden etabliertesten kognitiven Architekturen stellen hierbei Soar und ACT-R dar. In beiden Architekturen werden Modelle erstellt auf Basis der *rationalen Analyse Theorie* (Anderson 1991), nach der der Mensch versucht, für ein Ziel optimal zu handeln, indem er sich ein Abbild seiner Umgebung macht und möglichst wenig zusätzliche Annahmen treffen muss.

Soar basiert dabei auf der „Unified Theory of Cognition“, die den menschlichen Kognitionsprozess als Verarbeitung symbolischer Informationen beschreibt. Demnach werden Sensorinformationen in eine symbolische Ebene übersetzt. Auf dieser symbolischen Ebene stellt Soar ein Produktionssystem zur Verfügung, bei dem Sensoreingaben analysiert werden, um alle von einem Entwickler definierten Regeln zu finden, die anwendbar sind. Daraufhin wird eine dieser Regeln ausgelöst (Soar nennt dies ‚abfeuern‘), was den internen Wissensspeicher („working memory“) von Soar aktualisiert. Durch diese Aktualisierung werden neue Regeln anwendbar und der Prozess startet erneut, bis ein Zielzustand erreicht ist, keine anwendbaren Regeln mehr existieren oder das Abfeuern weiterer Regeln den Wissensspeicher nicht mehr verändert. In einem solchen *quiescence* genannten Zustand wird analysiert, welche Regeln die *Ausgabeverbinding* („output-link“) beeinflusst haben und dementsprechend eine Handlung erzeugt. Aufbauend auf Soar konnten (Brand und Schulte 2017) ein adaptives Automationssystem entwickeln, welches ein Verständnis der notwendigen Aufgaben mit den durchgeführten Aufgaben abgleicht und somit individuelle Automation für den Operateur eines simulierten Transporthelikopters generiert. Außerdem konnten (Ramaraj et al. 2020) in ihren Robotern ein Verständnis von menschlichen Absichten modellieren, sodass sie durch natürliche Interaktionsmuster mit einem menschlichen Lehrer Spiele und Puzzle erlernen können.

ACT-R legt hingegen einen Fokus auf die Begrenztheit des menschlichen Kurzzeitgedächtnisses und die Interaktionen damit. Die Modellierung des Frameworks beruht hierbei auf dem Verständnis von einzelnen Regionen des menschlichen Gehirns und welche dieser Regionen für welche Art von Aufgabe zuständig ist. Abbildung 10 stellt diese

Architektur dar. Die einzelnen Module können hierbei parallel arbeiten, aber innerhalb der Module wird sequenziell gearbeitet. Das Zusammenwirken der Module wird hierbei über das zentrale Element *Productions* geregelt, welches ähnlich wie in Soar Produktionsregeln enthält. Die Kommunikation läuft über Pufferspeicher, die immer nur ein einziges Element enthalten können. Die Produktionsregeln können hierbei auf alle Pufferspeicher zugreifen, um zutreffende Regeln zu aktivieren. Welche Regeln aktiviert werden, hängt vorwiegend von einer geschätzten *Nützlichkeit* („utility“) ab, die sich aus Schätzungen der Wahrscheinlichkeit, das Ziel mit dieser Regel zu erreichen, und der Kosten errechnet. Einen besonderen Stellenwert besitzt hierbei das *deklarative Modul* (oben rechts in Abbildung 10). Hierin können sogenannte *chunks* von Informationen abgespeichert und abgerufen werden. Wie schnell ein solches *chunk* abgerufen werden kann, hängt dabei von der Zeit ab, wann es das letzte Mal abgerufen wurde, sowie von der Häufigkeit des Abrufens. Hierdurch wird wiederum die *Nützlichkeit* der damit verbundenen Produktionsregeln und das Vergessen von Informationen kann modelliert werden. Durch diese Mechanismen ist es möglich Kognitionsmodelle des Menschen zu erstellen. Beispielsweise konnten (Taatgen und Lee 2003) ein Modell erzeugen, welches die Lernkurve der Probanden in einer vereinfachten Air Traffic Control Aufgabe beschreibt. Das Modell wurde mit den Daten aus dem ursprünglichen Experiment von (Ackerman 1988) verglichen und stellt eine Approximation dieser Daten dar. Aufgrund der Möglichkeit dadurch auch die menschlichen Fehler mit in die Modellierung mit einzubeziehen bezeichnen (Cassenti und Veksler 2017) ACT-R als die beste Option für kognitive Modellierung für adaptive Automation.

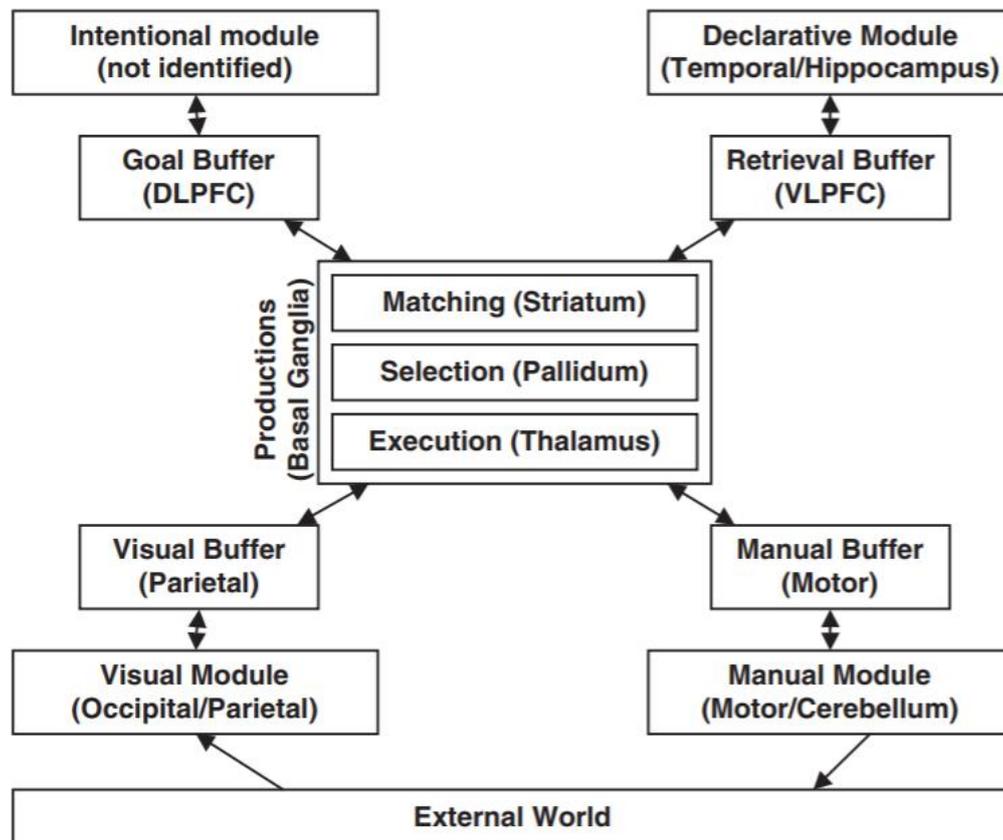


Abbildung 10: Überblick über ACT-R 6.0 aus (Taatgen et al. 2006)

3.1.2 VACP

Für eine Einschätzung der Arbeitsbeanspruchung der Piloten eines zukünftigen Helikopters haben (McCracken und Aldrich 1984) eine Analyse der durchzuführenden Aufgaben durchgeführt. Ziel dieser Analyse war es zu ermitteln, ob das Helikoptercockpit auf einen einsitzigen oder einen zweisitzigen Betrieb ausgelegt werden muss. Hierzu haben sie eine Metrik definiert, die die Aufgabendurchführung in die vier Kategorien Visuell, Auditiv, Kognitiv und Psychomotorisch (**VACP**) einteilt. In jeder dieser Kategorien wurden weiterhin sieben, beziehungsweise in der auditiven Kategorie vier, Schwierigkeitsstufen definiert. stellt die definierten Schwierigkeiten der einzelnen Kategorien dar.

Tabelle 1: Kategorien und Schwierigkeitswerte des VACP Modells (McCracken und Aldrich 1984)

Schwierigkeitswert	Beschreibung
	<u>Visuell</u>
1	Monitoren, Scannen, Beobachten
2	Erkennen von Bewegung oder Änderungen von Helligkeit oder Größe
3	Verfolgen oder Tracken
4	Zielen oder Ausrichten
5	Symbole, Zahlen oder Worte identifizieren
6	Aufgrund von mehreren Aspekten identifizieren
7	Lesen oder Text entschlüsseln
	<u>Auditiv</u>
1	Anwesenheit von Geräuschen erkennen
2	Veränderung der Lautstärke oder Tonhöhe erkennen
3	Inhalt einer Nachricht verstehen
4	Töne anhand von Signalmustern unterscheiden
	<u>Kognitiv</u>
1	Automatisch (unterbewusste Reaktion)
2	Zeichen/Signal wiedererkennen
3	Auswahl einer Alternative
4	Kodieren/Dekodieren oder Erinnern
5	Plan formulieren
6	Auswerten (Entscheidung aufgrund mehrerer Aspekte treffen)
7	Schätzen, Berechnen, Umrechnen
	<u>Psychomotorisch</u>
1	Diskrete Betätigung (z.B. Knopf)
2	Diskrete Anpassung (z.B. Drehrädchen)
3	Sprechen
4	Kontinuierliche Anpassung (z.B. Flight Controls)
5	Manipulationen (z.B. Karte bedienen)
6	Symbole produzieren (z.B. Schreiben)
7	Abfolge von Manipulationen (z.B. Tastatureingabe)

Im Folgenden wurde von Experten analysiert, welche Aufgaben in dem zukünftigen Helikoptercockpit sowohl im einsitzigen als auch dem zweisitzigen Fall von den Piloten durchgeführt werden müssen. Weiterhin wurde den Aufgaben eine zugehörige Schwierigkeit in den unterschiedlichen Kategorien zugeordnet und ermittelt, in welchen Missionsphasen

die Aufgabe wie lange durchgeführt werden muss. Schlussendlich wurden die einzelnen Missionsphasen in Zehn-Sekunden-Zeitfenster aufgeteilt und die Schwierigkeiten jedes dieser Zeitfenster wurden pro Kategorie aufaddiert. Jede Überschreitung des Maximalschwierigkeitswerts von sieben wurde hierbei als eine Überbeanspruchung gewertet.

Während sich der ursprüngliche Anwendungsfall auf einen Einsatz des Modells in der Entwurfsphase bezieht, können dieselben Kriterien auch in einer Echtzeitanalyse für adaptive Automation angewandt werden. Mit diesem Ansatz konnten (Rusnock und Geiger 2016) eine Erweiterung des VACP-Modells nutzen, um Schwellwerte zu identifizieren, bei denen eine adaptive Automation zusätzliche Unterstützungen aktivieren sollte. Die Erweiterung basiert auf dem VACP-Modell von (Mitchell 2000) und beinhaltet sieben statt der ursprünglichen vier Kategorien und erneut sieben Beanspruchungswerte in jeder Kategorie. Für die adaptive Automation wurde hierbei die Summe der Beanspruchungswerte der zu erledigenden Aufgaben errechnet. Überschreitet diese Summe innerhalb eines Zwei-Sekunden-Zeitfensters einen gewissen Schwellwert, werden zusätzliche Unterstützungsfunktionen aktiviert.

3.1.3 Single Resource Theory

In (Kahneman 1973) stellt Kahneman ein Modell zur Bewältigung von Aufgaben vor, bei dem die menschliche Aufgabenbewältigung auf eine einzelne Ressource, nämlich die Aufmerksamkeit, beschränkt wird. Diesem Ansatz liegt zugrunde, dass eine Aufgabe nicht durchgeführt werden kann, ohne dass der Mensch seine Aufmerksamkeit darauf richtet. Hierbei setzt Kahneman ‚Aufmerksamkeit ausrichten‘ mit ‚sich anstrengen‘ gleich. Somit ist der limitierende Faktor die Kapazität an zur Verfügung stehender Aufmerksamkeitsressourcen, welche von Situation zu Situation variabel sind. Diese Kapazität ist zum einen abhängig von dem Erregungszustand des Menschen und zum anderen von der ausgeübten Anstrengung, die beispielsweise durch Anreize wie Belohnungen stimuliert werden kann. In der in Abbildung 11 dargestellten, Visualisierung des Modells sind diese Einflussfaktoren auf die verfügbaren Aufmerksamkeitsressourcen mit ‚*Miscellaneous Determinants*‘ bezeichnet. Die zur Verfügung stehende Kapazität ist wiederum eines der Hauptaugenmerke für die ‚*Allocation Policy*‘. Hier findet auf Basis der derzeitigen Absichten (‚*Momentary Intentions*‘) die Auswahl statt, auf welche Aufgabe der Mensch seine Aufmerksamkeit richtet. Bei diesem Prozess werden die Absichten mit den Anforderungen der Aufgaben abgeglichen, um die Anstrengung zu ermitteln, die für die Aufgabenbewältigung notwendig ist. Dabei ist zu berücksichtigen, dass auch die Verarbeitung nicht relevanter Reize (‚*Enduring Dispositions*‘) eine Beanspruchung der Aufmerksamkeitsressourcen darstellt, für die Anstrengung aufgebracht werden muss. Dieses

Aufbringen von Anstrengung modifiziert die Grenze der verfügbaren Kapazitäten und führt gleichzeitig dazu, dass Manifestationen von Erregung (*Miscellaneous Manifestations of Arousal*) ausgelöst werden. Die Manifestationen bilden die Grundlage der Messung von psychophysiologischen Parametern, wie sie in Kapitel 3.2.1 betrachtet werden.

Somit wird laut Kahnemann Aufgabendurchführung möglich, indem der Mensch seine Aufmerksamkeit auf eine Aufgabe richtet und je nach Schwierigkeit zusätzliche Anstrengung aufbaut, um die Menge an verfügbaren Aufmerksamkeitsressourcen zu erhöhen. Wichtig ist hierbei zu erwähnen, dass Aufmerksamkeit auf mehrere Aufgaben verteilt werden kann und bestimmt werden muss, welcher Aufgabe wie viele der Aufmerksamkeitsressourcen zuteilwerden.

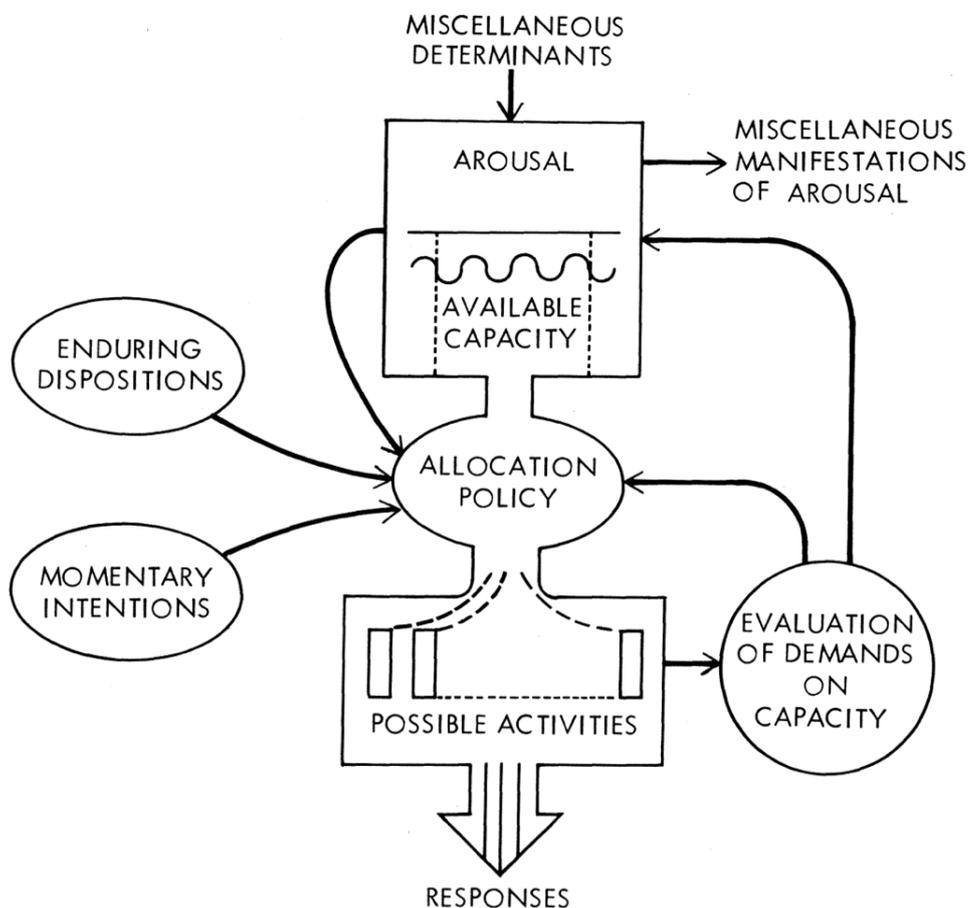


Abbildung 11: Darstellung der Single Resource Theory nach (Kahneman 1973)

Die Nutzung dieses Modells innerhalb eines adaptiven Automationsystems würde auf der Beschreibung der notwendigen Ressourcen zur Bewältigung der einzelnen Aufgaben beruhen. Gemeinsam mit einer Erkennung, welche Aufgabe vom Menschen bearbeitet wird, kann hierdurch ein Modell erstellt werden, wie viele Aufmerksamkeitsressourcen benötigt werden. Ein Abgleich der benötigten Ressourcen mit einer Vorstellung der tatsächlich vorhandenen Ressourcen beschreibt die Notwendigkeit von automatisierten

Unterstützungsfunktionen. Allerdings konnte kein Einsatz des Modells in einem adaptiven Assistenzsystem gefunden werden. Einzig die Nutzung des Modells zur Berücksichtigung der verfügbaren mentalen Ressourcen bei der Planerstellung von (Heilemann und Hollatz 2020) nutzt das beschriebene Modell innerhalb eines automatisierten Systems. Bei diesem Planungsansatz werden zusätzliche Abhängigkeiten zwischen Aufgaben innerhalb eines Planungsproblems genutzt, um eine Überforderung des menschlichen Operateurs bereits in der Planung einer militärischen Mission zu vermeiden.

3.1.4 Multiple Resource Theory

Ein weit verbreitetes mentales Modell der Beanspruchung eines menschlichen Operateurs basiert auf der ‚Multiple Resource Theory‘ von (Wickens 2002). Diese Theorie teilt die Verarbeitung von Informationen in die vier orthogonalen Dimensionen ‚Code‘, ‚Modalities‘, ‚Stages‘ und ‚Visual Processing‘. Jede Dimension stellt eine Unterscheidung der Art der zu verarbeitenden Information dar. Die Dimension Code differenziert hierbei zwischen räumlich analogen, wie z.B. die Entfernung eines Objekts, und verbalen Informationen, wie die Kategorie eines Objekts. Bei der Dimension Modalities handelt es sich um die Unterscheidung zwischen den Wahrnehmungsmodalitäten visuell und auditiv. Die visuelle Wahrnehmung beinhaltet die Dimension Visual Processing, welche in fokale und periphere Sicht unterteilt ist. Die letzte Dimension – Stages - trennt die unterschiedlichen Phasen des Wahrnehmungsprozesses voneinander ab. Hierbei wird zwischen Wahrnehmung, Verständnis und Reaktion unterschieden.

Diese Aufteilung spiegelt wider, welcher Teil des menschlichen Gehirns beansprucht wird, wenn eine entsprechende Aufgabe ausgeführt wird. Dementsprechend modelliert die ‚Multiple Resource Theory‘ das Konfliktpotential einzelner Aufgaben zueinander. Wie in Abbildung 12 dargestellt ist zu beachten, dass in den Stages Verständnis und Reaktion die Modalitätsdimension nicht unterschieden wird.

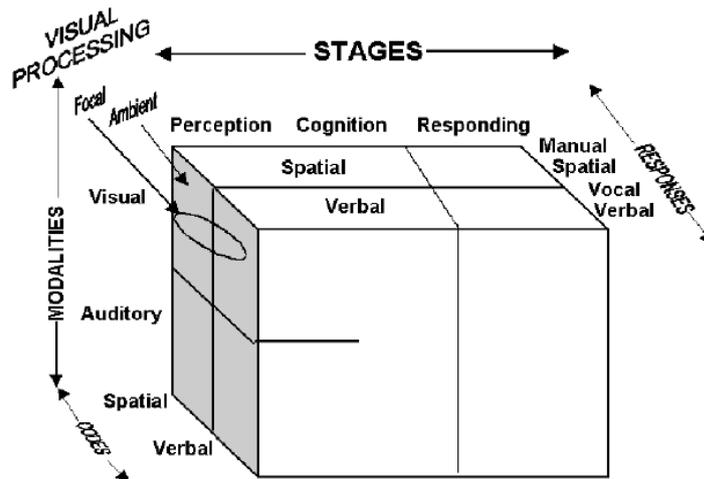


Abbildung 12: Dreidimensionale Darstellung der ‚Multiple Resource Theory‘ nach (Wickens 2002)

Werden in einer Situation mehrere Aufgaben gleichzeitig ausgeführt, kann hiermit ein Beanspruchungswert für den Konflikt zwischen den Aufgaben errechnet werden. Dieser Ansatz wurde bereits 1989 in (North und Riley 1989) unter dem Namen W/INDEX veröffentlicht. Nach (Wickens 2002) wird hierbei jedem Ressourcenpaar ein eindeutiger Konfliktwert zugeordnet wie in Tabelle 2 dargestellt. In dieser Konfliktmatrix werden die Probleme gleichartiger Ressourcenbeanspruchungen noch einmal genauer ausdefiniert. Beispielsweise besteht ein großer Ressourcenkonflikt zwischen zwei Aufgaben, die eine manuelle Antwort benötigen, als bei zwei Aufgaben, die in einer anderen Ressource übereinstimmen.

Tabelle 2: Konflikte der Ressourcenpaare nach (Wickens 2002)

		Ressourcen von Aufgabe A							
		VS	VV	AS	AV	CS	CV	RS	RV
Ressourcen von Aufgabe B	VS	0,8	0,6	0,6	0,4	0,7	0,5	0,4	0,2
	VV		0,8	0,4	0,6	0,5	0,7	0,2	0,4
	AS			0,8	0,4	0,7	0,5	0,4	0,2
	AV				0,8	0,5	0,7	0,2	0,4
	CS					0,8	0,6	0,6	0,4
	CV						0,8	0,4	0,6
	RS							0,8	0,6
	RV								1,0

Innerhalb von adaptiven Automationssystemen wurde dieses Modell bereits mehrfach erfolgreich eingesetzt. So konnte (Maiwald 2015) ein Echtzeitmodell der durchgeführten Aufgaben in einer militärischen Helikoptermission nutzen, um eine Abschätzung der beanspruchten Ressourcenkanäle zu generieren. Mithilfe dieser Ressourcenschätzung hat er einen passenden Ressourcenkanal gewählt, um möglichst intrusionsfrei Unterstützungsfunktionen bereitstellen zu können. Die resultierende Automation sorgte für eine verbesserte Leistung bei reduzierter Beanspruchung der Versuchsteilnehmer im Vergleich zu einer nichtadaptiven Bereitstellung der Unterstützungsfunktionen.

Weiterhin haben (Schmitt et al. 2018) diesen Ansatz erweitert, indem sie eine detailliertere Aufgabenerkennung mit einer Planverständniskomponente kombiniert haben. Diese Kombination erlaubt es ihrem System, die Aufgabensituation genauer zu verstehen und durch einen Abgleich zwischen notwendigen und durchgeführten Aufgaben sogar zukünftige Überbeanspruchungssituationen vorherzusehen und diese proaktiv zu vermeiden.

3.1.5 Cognitive Reliability and Error Analysis Method

Basierend auf dem *Context and Control* (Hollnagel 1993) Modell, bei dem der menschliche Kognitionsprozess vorwiegend als ein sich wiederholender Prozess aus Beobachten und Handeln modelliert wird, hat (Hollnagel 1998) die *Cognitive Reliability and Error Analysis Method* (CREAM) vorgestellt. Bei dieser Methode wird die Beanspruchung des Operators durch die Arbeitsbedingungen („*Common Performance Conditions* (CPC)“) der durchzuführenden Aufgabe modelliert, um Ursachen menschlicher Fehler analysieren und vorhersagen zu können. Als Arbeitsbedingungen werden hier einerseits grundlegende Faktoren, wie die Tageszeit oder die Eignung der Mensch-Maschine-Schnittstelle, berücksichtigt, andererseits aber auch individuelle Faktoren wie der Trainingszustand des Operators.

Das Modell beinhaltet einen zweistufigen Prozess aus einer vereinfachten Analyse, um Aufgaben mit erhöhtem Fehlerrisiko zu identifizieren und einer weiterführenden Analyse für Aufgaben, bei denen ein hohes Fehlerrisiko festgestellt wurde, um dieses zu quantifizieren und Möglichkeiten zur Fehlerminimierung zu finden.

Die Prozesstufe 1 (*Basic Methode*) beginnt mit einer Beschreibung der durchzuführenden Aufgaben in Form einer Aufgabenanalyse wie beispielsweise der „*Hierarchical Task Analysis*“ (Annett 2003). Diese erlaubt es die Aufgaben in eine zeitliche Ablaufstruktur zu überführen. Daraufhin werden die „*Common Performance Conditions*“ untersucht, wobei jeder Arbeitsschritt der Aufgabe hinsichtlich der Arbeitsbedingungen genauer betrachtet wird. Das Modell unterscheidet hierbei neun Kategorien von Arbeitsbedingungen und für jede der Kategorien wird eine Aussage abgegeben, ob sie die Zuverlässigkeit der

menschlichen Leistung erhöht, senkt oder unbeeinflusst lässt. Anschließend werden die getroffenen Aussagen über die Arbeitsbedingungen unabhängig voneinander aufsummiert und ein Triplet erzeugt. Diese besteht aus der Summe an Bedingungen, welche die (1) Zuverlässigkeit erhöhen, diese (2) senken oder keinen (3) Einfluss haben. Hierbei ist zu beachten, dass nicht alle Arbeitsbedingungen einen positiven Einfluss auf die Zuverlässigkeit haben können. So kann beispielsweise die Anzahl gleichzeitig verfolgter Arbeitsziele nur einen negativen oder keinen Einfluss haben. Kein Einfluss tritt ein, wenn die Anzahl gleichzeitig verfolgter Arbeitsziele niedrig ist und gleichermaßen ein negativer Einfluss, wenn sie hoch ist. Das somit erzeugte Triplet kann, wie in Abbildung 13 zu sehen ist, in vier Kontrollmodi eingeteilt werden. Dazu wird die Summe der Zuverlässigkeitssenkungen auf der horizontalen Achse und die Summe der Erhöhungen auf der vertikalen Achse aufgetragen und der zugehörige Kontrollmodus abgelesen. Der letzte Schritt der *Basic* Methode ist eine Zuordnung des ermittelten Kontrollmodus zu einem Fehlerwahrscheinlichkeitsbereich. Hierzu stellt Hollnagel eine statische Zuordnung vor, bei der beispielsweise der „*opportunistische*“ Kontrollmodus einem Bereich von $10E-2$ bis $0.5E-0$ Fehlern pro Ausführung zugeordnet wird.

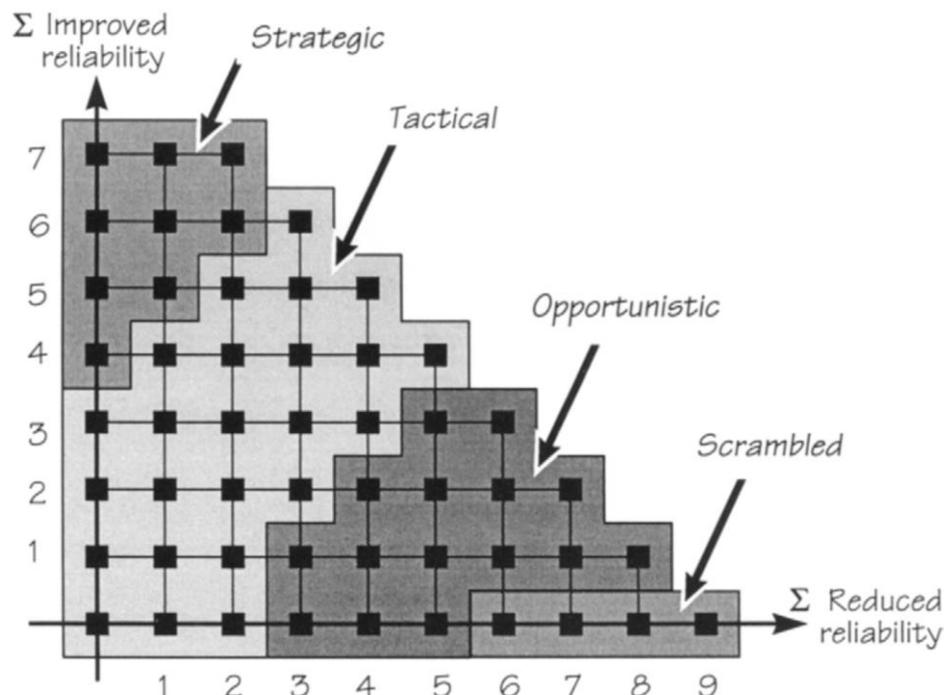


Abbildung 13: Zusammenhang zwischen CPC Triplet und dem Kontrollmodus

Die Prozessstufe 2 (*Extended* Methode) erlaubt eine genauere Abschätzung der Fehlerwahrscheinlichkeit, aufbauend auf den Erkenntnissen der *Basic* Methode. Dazu

werden die einzelnen Arbeitsschritte, die in der *Hierarchical Task Analysis* identifiziert wurden, einer zugehörigen dominanten kognitiven Aktivität zugeordnet. Bei diesen kognitiven Aktivitäten handelt es sich um allgemeine Aufgabenklassen wie *Vergleichen*, *Beobachten* oder *Ausführen*. Infolgedessen kann anhand von Abbildung 14 abgelesen werden, welche kognitiven Grundfunktionen des *COCOM-Modells* (*Observation*, *Interpretation*, *Planning* und *Execution*) für die Aktivität benötigt werden. Die kognitiven Grundfunktionen entsprechen hierbei den vier Stufen der menschlichen Informationsverarbeitung nach (Parasuraman et al. 2000) (siehe Kapitel 2.2). Die Bestimmung der kognitiven Grundfunktionen kann optional genutzt werden, um ein „kognitives Beanspruchungsprofil“ der Aufgabe zu erstellen, indem die Vorkommnisse der Grundfunktionen in den Aufgabenschritten aufsummiert werden. Außerdem werden mit Hilfe des *COCOM-Modells* die kognitiven Aktivitäten genutzt, um die Wahrscheinlichkeit des wahrscheinlichsten Fehlers zu bestimmen. Abschließend wird der Kontrollmodus genutzt, um die resultierenden Fehlerwahrscheinlichkeiten noch einmal zu adaptieren, indem diese mit einem Gewichtungsfaktor für den jeweiligen Kontrollmodus multipliziert wird.

Activity type	COCOM function			
	Observation	Interpretation	Planning	Execution
Co-ordinate			♦	♦
Communicate				♦
Compare		♦		
Diagnose		♦	♦	
Evaluate		♦	♦	
Execute				♦
Identify		♦		
Maintain			♦	♦
Monitor	♦	♦		
Observe	♦			
Plan			♦	
Record		♦		♦
Regulate	♦			♦
Scan	♦			
Verify	♦	♦		

Abbildung 14: Zuordnung von Aktivitäten zu benötigten kognitiven Grundfunktionen nach (Hollnagel 1998)

Für den Einsatz in automatisierten Systemen ist die *Cognitive Reliability and Error Analysis Method*, sowie deren Nachfolger „*The Functional Resonance Analysis Method*“ (Patriarca et al. 2020) laut des Autors nicht geeignet, da sie die manuelle Analyse der Aufgaben und Arbeitsbedingungen benötigt und somit nicht automatisiert durchgeführt werden kann. Im Gegensatz dazu ist die Methode gut geeignet, um Aufgaben zu identifizieren, bei denen

automatisierte Unterstützungsfunktionen notwendig sind, um eine fehlerfreie Aufgabendurchführung zu begünstigen.

3.1.6 Cognitive Task Load Modell

Das *Cognitive Task Load* (CTL) Modell von (Neerinx und van Besouw 2001) bildet die kognitive Beanspruchung auf drei Dimensionen ab, wodurch die Notwendigkeit von kognitiver Unterstützung ermittelt wird. Die erste Dimension beschreibt die Zeit, in der der Mensch mit der Aufgabenbearbeitung beschäftigt ist. Die Autoren beschreiben, dass laut (Beevis et al. 1994) diese Zeit eine Grenze von 70 bis 80 Prozent überschreiten sollte, um geeignete Beanspruchung zu garantieren. Die zweite Dimension ist der Grad der, zur Aufgabenbearbeitung notwendigen, Informationsverarbeitung. Hierbei wird sich auf das Rasmussen Schema aus Kapitel 2.2 bezogen. Demnach sind „*skill-based*“ Aufgaben sehr leicht zu bearbeiten und „*knowledge-based*“ Aufgaben haben einen hohen Grad an notwendiger Informationsverarbeitung. Wichtig hierbei ist neben der durchzuführenden Aufgabe auch die Darstellung der Informationen. Somit können leichte Aufgaben einen hohen Grad an Informationsverarbeitung benötigen, wenn die dazugehörige Benutzerschnittstelle die Aufgabe nicht passend darstellt. Dies wird im CTL Modell explizit berücksichtigt. Die letzte Dimension bezieht sich auf die Anzahl an unabhängigen Aufgabenkontexten, die mit der übergeordneten Aufgabe verknüpft sind. In diesem Zusammenhang bedeuten Aufgabenkontexte, einfache Handlungsabfolgen („*Basic Action Sequences*“), wie beispielsweise das Abarbeiten einer Checkliste, um ein gezieltes Problem zu beheben.

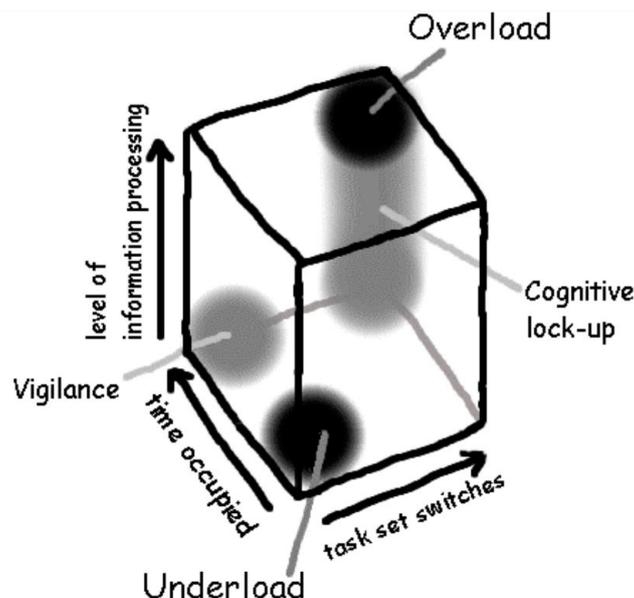


Abbildung 15: Kritische Bereiche im CTL Modell (entnommen aus (Neerinx und van Besouw 2001))

Abbildung 15 beschreibt die Aufgabensituationen innerhalb der Dimensionen, in denen kritische Beanspruchungsniveaus entstehen können. Im Falle von niedrigen Bewertungen in allen Dimensionen besteht die Gefahr der Unterforderung („*Underload*“). Für jeden Beanspruchungszustand, bewertet das Modell außerdem die Zeit, die der Mensch darin verbringt. Hiernach ist die Unterforderungssituation erst nach einer Dauer von 5 Minuten als kritisch zu bewerten.

Ein weiterer kritischer Bereich ist charakterisiert durch eine langanhaltende Aufgabe, die wenig Informationsverarbeitung, aber für mehr als 80% der Zeit die Aufmerksamkeit des Operateurs benötigt. Innerhalb des CTL Modells treten hierbei Vigilanz Dekremente auf, wenn eine solche Aufgabensituation für mehr als 20 Minuten anhält. Eine verringerte Vigilanz führt dabei zu „*Out-of-the-loop performance*“ (siehe 2.4.2) oder „*Complacency*“ (siehe 2.4.3) Problemen.

Unabhängig von dem Grad der Informationsverarbeitung besteht bei einem hohen Beschäftigungsanteil und einer hohen Anzahl an Kontextwechseln das Problem des „*Cognitive Lock-up*“. Bei diesem Zustand werden Aufgaben sequenziell anstatt parallel abgearbeitet (Kerstholt und Passenier 2000), weil Menschen dazu neigen Aufgaben abzuarbeiten, bevor sie eine neue Aufgabe beginnen (Boehne und Paese 2000).

Die letzte kritische Aufgabensituation stellt eine hohe Beanspruchungsbewertung in allen drei Dimensionen dar. Dieser Zustand beschreibt eine Überbeanspruchungssituation („*Overload*“) also ein Überschreiten der „*red-line of workload*“ aus Kapitel 2.5.2.7. Die Folgen von Überbeanspruchung können erneut den „*Cognitive Lock-up*“ beinhalten, aber auch Verhaltensänderungen wie das bewusste Ignorieren von Teilaufgaben oder das Beenden der Aufgabenbearbeitung.

Zur Bewertung der einzelnen Dimensionen nutzen die Autoren eine Aufgabenanalyse auf Basis von einzelnen Anwendungsfällen. Hierbei werden Experten in der Aufgabendomäne befragt, um kritische Anwendungsfälle zu identifizieren und die notwendigen Schritte zur Bearbeitung des Anwendungsfalls zu beschreiben. Mit Hilfe dieser Analyse können kritische Beanspruchungssituationen ermittelt und die dazugehörigen Auslöser adressiert werden. Eine vorgestellte Möglichkeit, eine Überbeanspruchungssituation zu lindern ist das Einführen einer Prioritätenliste für vorhandene Warnungen und dadurch die Anzahl an Aufgabenkontextwechseln zu verringern.

Das Modell wird vorwiegend zum Zweck der Analyse von Beanspruchungssituationen genutzt, um Möglichkeiten aufzudecken, die kognitive Unterstützung liefern um kritische Situationen bereits in der Gestaltung der Aufgabe zu vermeiden. Allerdings können die beschriebenen drei Dimensionen mithilfe von automatisierter Aufgabenermittlung in Echtzeitanwendungen bewertet werden und dadurch der Bedarf an kognitiven

Unterstützungsfunktionen während der Aufgabendurchführung ermittelt werden. Mit diesem Ansatz wird in (Schwarz et al. 2019) ein Auslöser für adaptive Anpassung der Automation generiert, der im Rahmen eines größeren Frameworks (RASMUS) die adaptive Automation steuert.

3.1.7 NASA-TLX

Der NASA Task Load Index (Hart und Staveland 1988) ist ein Fragebogen zur Selbsteinschätzung der benötigten Beanspruchung innerhalb eines Experiments oder der Durchführung einer Arbeit. Innerhalb des Fragebogens muss beantwortet werden, wie hoch die Beanspruchung innerhalb der vergangenen Aufgabendurchführung war, wobei die Beanspruchung in sechs Unterdimensionen aufgeteilt wird. Die abgefragten Dimensionen sind die mentale sowie die körperliche Anforderung, die zeitliche Anforderung, die Güte der erbrachten Leistung, die notwendige Anstrengung und die Frustration während der Aufgabendurchführung. Innerhalb jeder dieser Dimensionen muss die Beanspruchung auf einer Skala von 1 bis 20 bewertet werden. Zusätzlich beinhaltet der Fragebogen eine paarweise Gewichtung der Beanspruchungsdimension, um weiterhin zu bewerten, welche der Dimensionen für den Versuchsteilnehmer die Beanspruchung am meisten beeinflusst hat. In einem Review über die Methode (Hart 2006) beschreiben die ursprünglichen Autoren allerdings, dass die Gewichtung der einzelnen Dimensionen selten durchgeführt und ausschließlich die Bewertung der sechs beschriebenen Beanspruchungsdimensionen genutzt wird. Diese vereinfachte Form der Methode wird als „*Raw TLX*“ bezeichnet und bietet laut (Bustamante und Spain 2008) sogar ein genaueres Abbild der tatsächlichen Beanspruchung, da die Gewichtung der Dimensionen von den Versuchsteilnehmern häufig nicht verstanden wird.

Der Einsatz dieser Methode innerhalb eines automatisierten Systems ist aufgrund ihrer Natur als Fragebogenmethode nahezu unmöglich. Dennoch bietet sie seit der ursprünglichen Veröffentlichung die Möglichkeit, ein Abbild der Beanspruchung als Referenzwert zu generieren. Hierbei werden häufig sogenannte „*freeze probe*“ Techniken eingesetzt, in denen die Simulation innerhalb eines Experiments angehalten wird, um rückwirkend die Beanspruchung bewerten zu lassen. Die gesammelten subjektiven Einschätzungen können dann genutzt werden, um sie mit den Ergebnissen von beispielsweise Echtzeitmessmethoden zu vergleichen und diese somit neu auszurichten.

3.2 Messung von Beanspruchungsindikatoren

Bei der Messung von Beanspruchungsindikatoren werden bekannte Zusammenhänge zwischen messbaren Größen und der Beanspruchung genutzt, um aus einer Veränderung der

Messgröße eine Veränderung der Beanspruchung abzuleiten. Die dabei bekannten Messgrößen lassen sich dabei in die zwei Kategorien der kontextfreien- und der kontextbehafteten Messgrößen unterteilen. Die kontextfreien Messgrößen werden durch übertragbare Messungen abgebildet, bei denen Reaktionen des menschlichen Körpers direkt mit der Beanspruchung verknüpft werden. Eine große Schwierigkeit bei diesem Ansatz ist die Unterscheidung zwischen Veränderungen der Messgröße aufgrund einer Beanspruchungsänderung und einer beanspruchungsunabhängigen Änderung der Messgröße. Demgegenüber stehen die kontextbehafteten Messgrößen bei denen Domänenwissen genutzt wird um die Modelle aus Kapitel 3.1 in einem Echtzeitkontext anwenden zu können.

3.2.1 Kontextfreie Ermittlungsmethoden

Bei den kontextfreien Methoden zur Beanspruchungsermittlung handelt es sich um Methoden, für deren Einsatz kein domänenspezifisches Wissen notwendig ist. Der Vorteil dieser Methoden besteht darin, dass sie grundlegende Zusammenhänge von Parametern wie Beanspruchung, Erregung und Anstrengung mit Veränderungen im menschlichen Körper abbilden. (Singh 2021) stellt die dabei geläufigen Messmethoden vor, die sich grundlegend auf das Herz, das Gehirn oder das Auge beziehen. Zusätzlich werden abstraktere Messmethoden genutzt, wie die Komplexität der verwendeten Sprache oder eine Erkennung der Emotionen.

3.2.1.1 Herz

Für die Schätzung der Beanspruchung über eine Messung der kardialen Aktivität wird ein Elektrokardiograph (EKG) eingesetzt. Dieser misst den elektrischen Impuls der die Kontraktion des Herzmuskels verursacht. Dieser QRS-Komplex genannte Impuls besteht aus drei Zacken, die nacheinander erfolgen. Der in Abbildung 16 (rechts) dargestellte Komplex beginnt mit einem negativen Ausschlag der Q-Zacke genannt wird und die Depolarisation des Septums anzeigt. Darauf folgt mit der R-Zacke ein starker positiver Ausschlag, der die Aktivität der linken Herzkammer darstellt. Demgegenüber steht die kleinere negative S-Zacke der rechten Herzkammer. Für einen kompletten Herzschlag ist der QRS-Komplex von der P-Welle und der T-Welle umgeben. Für die Messung werden drei Dioden wie in Abbildung 16 (links) am Körper platziert. Dabei wird jeweils ein Sensor leicht unter dem Schlüsselbein und ein dritter Sensor direkt unterhalb der untersten linken Rippe angebracht (Fortin-Côté et al. 2019).

Einer der in diesem Rahmen gut untersuchten Zusammenhänge ist der zwischen körperlicher Anstrengung und der Herzrate (Nassef et al. 2010). Dieser Zusammenhang wird inzwischen

in Alltagsgeräten genutzt, um die körperliche Anstrengung mitzuschneiden und dadurch Analysen über die Länge der sportlichen Aktivität (hohe körperliche Anstrengung) oder des Schlafs der Nutzer (niedrige körperliche Anstrengung) zu erzeugen.

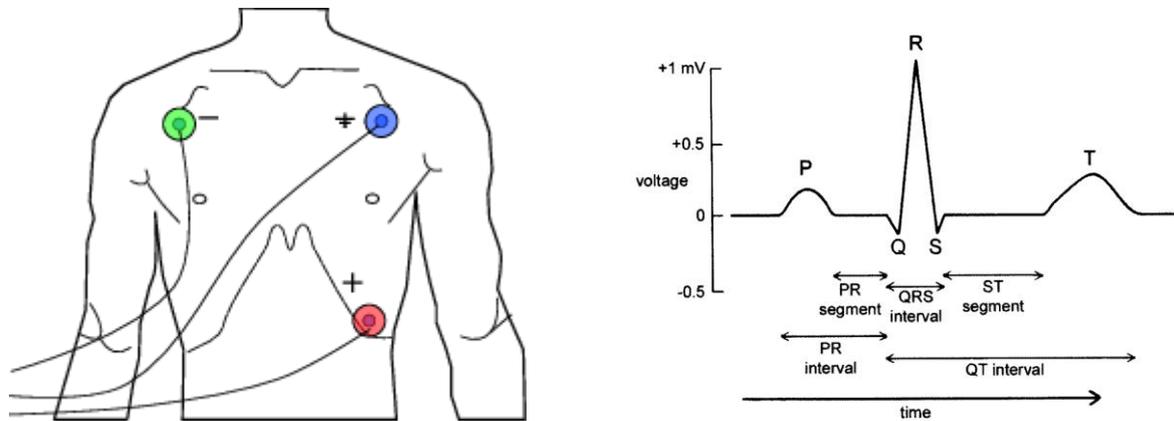


Abbildung 16: (links) Diodenplatzierung für die EKG-Messung nach (Fortin-Côté et al. 2019) (rechts) Eine typische EKG-Welle nach (Dupre et al. 2005)

Neben der körperlichen Anstrengung können mit Hilfe der EKG-Messung allerdings auch Indikatoren für mentale Beanspruchung ermittelt werden. So konnten (Dehais et al. 2012) und (Vuksanović und Gal 2007) eine steigende Herzrate, sowie eine sinkende Herzratenvariabilität mit steigender mentaler Beanspruchung nachweisen. Die Herzrate ist dabei Anzahl der Herzschläge pro Minute und die Herzratenvariabilität ist die Varianz der zeitlichen Abstände aufeinanderfolgender Schläge. Weitere zeitbasierte Methoden sind die ‚*root mean square of successive differences*‘ (RMSSD) (Lyu et al. 2017; Mehler et al. 2011) und das Integral eines Histogramms der RR-Intervalle geteilt durch seine Höhe (RR-Tri) (Electrophysiology, Task Force of the European Society of Cardiology the North American Society of Pacing 1996), die beide sinken, wenn die Beanspruchung steigt.

Weitere Beanspruchungsindikatoren können aus einer Poincaré-Analyse oder einer Analyse des Frequenzspektrums extrahiert werden. Bei der Poincaré Analyse sinkt die Standardabweichung der zweiten Hauptachse (Poincaré SD2) (Mukherjee et al. 2011), während bei der Analyse des Frequenzspektrums der Anteil an hochfrequenten (0,15 Hz – 0,4 Hz) Signalanteilen abnimmt (Cinaz et al. 2010). Allerdings benötigen diese Methoden einen Zeitraum von 5 Minuten an Datenpunkten, damit zuverlässige Aussagen getroffen werden können (Li et al. 2019). Somit ist diese Art von Analyse für Echtzeitanalysen ungeeignet. Tabelle 3 fasst die Methoden noch einmal zusammen.

Tabelle 3: Herzbasierte Methoden zur Beanspruchungsermittlung

Messwert	Einheit	Beschreibung	Erwartete Änderung
Herzrate	1/min	Die durchschnittliche Anzahl an Herzschlägen pro Minute	Anstieg
RMSSD	ms	Quadratwurzel der quadrierten Abweichungen aufeinanderfolgender RR Intervalle	Senkung
RRTri	-	Integral des RR-Interval Histograms geteilt durch seine Höhe	Senkung
HF normalized	-	Der normalisierte Hochfrequenzanteil (0.15-0.4Hz) der Herzratenvariabilität	Senkung
Poincaré SD2	ms	Standardabweichung der zweiten Hauptachse der Poincaré Analyse	Senkung

3.2.1.2 Gehirn

Die automatisierte Messung der Gehirnaktivität stellt grundsätzlich einen Trade-off zwischen einer räumlichen Auflösung, also der Genauigkeit der Bestimmung der aktiven Hirnregionen, und der zeitlichen Auflösung dar (Parasuraman und Rizzo 2006). Eine hohe zeitliche Auflösung bei einer niedrigen räumlichen Auflösung liefern die Elektroenzephalographie (EEG) und die funktionelle Nahinfrarot-Spektroskopie (fNIRS). Demgegenüber liefert die funktionelle Magnetresonanztomographie (fMRI) eine hohe räumliche Auflösung kann aber keine hohe zeitliche Auflösung erreichen, wodurch die Methode wenig geeignet für adaptive Systeme ist. Weitere Methoden, wie die Magnetoenzephalographie (MEG) und die Elektrokortikographie (ECoG) können genauere Ergebnisse erzeugen, sind allerdings bisher nicht für den Einsatz außerhalb von medizinischen Anwendungen geeignet, da sie entweder große Laboraufbauten benötigen (MEG) oder die Sensoren direkt auf das Gehirn aufgelegt werden müssen (ECoG).

Elektroenzephalographie

Bei einem EEG werden die Ströme gemessen, die aufgrund von Hirnaktivitäten auftreten. Dazu werden zwischen 8 und 32 Elektroden auf dem Kopf platziert. Damit die Elektroden an den richtigen Stellen angebracht werden sind die für gewöhnlich an einer Kappe angebracht, die dem Probanden aufgesetzt wird. Die gemessenen Ströme werden in mehrere Frequenzbänder aufgeteilt, die entsprechenden Aktivitäten zugeordnet werden können (Brouwer et al. 2012).

Das Alpha Frequenzband (8-12 Hz) beschreibt den Grundrhythmus des Gehirns. Eine hohe Aktivität dieses Frequenzbands signalisiert einen Ruhezustand und mit steigender Anstrengung sinkt die Aktivität.

Das Beta Frequenzband (13-30 Hz) wird bei hohen Stresssituationen oder Angespanntheit aktiviert. Zusätzlich signalisiert dieses Frequenzband logisches Denken oder bewusste Überlegungen.

Das Gamma Frequenzband (>30 Hz) signalisiert Aufmerksamkeit und die Aktivität des Gedächtnisses.

Das Theta Frequenzband (4-8 Hz) zeigt Aktivitäten bei kognitiver Anstrengung, aber auch bei emotionalen Reaktionen. Gleichzeitig sinkt die die Aktivität in diesem Frequenzband bei Müdigkeit.

Das Delta Frequenzband (1-4 Hz) ist vorwiegend beim Tiefschlaf aktiv.

Nach (Brouwer et al. 2012) stellen das Alpha- und das Theta Frequenzband die geeignetsten Indikatoren für Beanspruchung dar. Eine hohe Alpha Aktivität signalisiert einen Ruhezustand und somit eine niedrige Beanspruchung. Das Thetaband dient nach (Borghini et al. 2012) zusätzlich als Indikator für den Zugriff auf das Arbeitsgedächtnis und wird zusätzlich aktiv, wenn eine mentale Fokussierung aufrechterhalten wird (Gevins und Smith 2003).

Neben der reinen Frequenzanalyse können auch zeitliche Analysen des EEG Signals genutzt werden, bei denen die Aktivität des Signals mit auftretenden Ereignissen in Verbindung gebracht wird (Fu und Parasuraman 2006). Ein probates Maß für diese „*Event Related Potentials*“ stellt das P300-Maß, das einen positiven Ausschlag 300ms nach dem Auftreten eines Ereignisses ermittelt. Dieser Ausschlag ist größer, wenn die Aufmerksamkeit auf das Ereignis gerichtet wird. Zusätzlich sinkt die Amplitude des Ausschlags mit einer erhöhten Beanspruchung (Friedrich et al. 2011). Laut (Roy et al. 2016) ist die Analyse der *Event Related Potentials* robuster gegen Müdigkeit und Vigilanz Probleme als die Spektralanalyse. Allerdings wird hierfür ein Verständnis der Domäne benötigt um Ereignisse zu identifizieren, weshalb dies eher eine kontextbehaftete Messung darstellt.

Funktionelle Nahinfrarot-Spektroskopie

Bei der fNIRS wird die Veränderung der Sauerstoffsättigung des Bluts in unterschiedlichen Gehirnregionen gemessen, die durch Gehirnaktivitäten ausgelöst wird (Herff et al. 2014). Bei mental aufwändigen Aufgaben wird der präfrontale Cortex aktiv und benötigt deshalb mehr Sauerstoff, weshalb dieser mit mehr sauerstoffhaltigem Blut versorgt wird. Daher steigt der Anteil an sauerstoffgesättigtem Hämoglobin im präfrontalen Cortex laut (Causse et al. 2017) mit steigender Aufgabenkomplexität, während der Anteil an ungesättigtem Hämoglobin sinkt. Zusätzlich konnten sie aus einer Kombination dieser fNIRS Messwerte und der Leistung der Versuchsteilnehmer ein Maß für die Effizienz erzeugen.

3.2.1.3 Auge

Eine weitere Form Beanspruchung zu ermitteln, stellen okulomotorische Messungen dar. Hierunter fallen alle Parameter, die sich auf das menschliche Auge beziehen. Die Vermessung des Auges geschieht dabei für gewöhnlich über kamerabasierte Messverfahren bei denen Kameras an der Mensch-Maschine Schnittstelle positioniert werden, damit sie die Kopfposition sowie das Auge erfassen können. Die Kameras sollten dabei im unteren Bereich des Sichtfelds positioniert werden, damit die Pupille möglichst jederzeit sichtbar ist. Mit Hilfe von einer Modellierung der Augengeometrie kann das Auge in den Kamerabildern identifiziert und von Bild zu Bild nachverfolgt werden (Nitschke et al. 2013). Für die sogenannte Cornea-Reflex Methode werden Infrarotkameras eingesetzt und mit Infrarotstrahlern kombiniert. Dies erlaubt zusätzlich eine Bestimmung der Blickrichtung, indem das Licht des Infrarotstrahlers auf der Hornhaut des Auges (Cornea) reflektiert wird und in Relation zu der Iris gesetzt werden kann (siehe Abbildung 17). Ein 3D-Modell der Infrarotstrahler und den Elementen der Mensch-Maschine Schnittstelle erlaubt es die bestimmte Blickrichtung in Koordinaten auf den Anzeigeelementen zu übersetzen.



Abbildung 17: Blickrichtungsbestimmung mit Cornea-Reflex Methode. Reflektion des Infrarotstrahlers (grüne Punkte); Positionsschätzung der Iris (rote Kreise) und geschätzte Blickrichtung (grüne Linie) entnommen aus (Nitschke et al. 2013)

Aus der Vermessung des Auges können einige statistische Maße extrahiert werden, die als Indikatoren für Beanspruchung dienen können. Die geläufigen Parameter in der Literatur sind hierbei die folgenden:

Blinzelfrequenz

Mit der Blinzelfrequenz wird gemessen, wie häufig in einer Minute geblinzelt wird. Die Blinzelfrequenz steigt, bei einer kognitiven Nebenaufgabe, während sie andererseits sinkt, wenn die Schwierigkeit in der Hauptaufgabe steigt (Faure et al. 2016).

Blinzeldauer

Die Blinzeldauer beschreibt die Länge eines einzelnen Blinzeln, also die Zeit, die das Augenlid geschlossen ist. Mit einer steigenden Beanspruchung wird die Länge des Blinzeln reduziert, damit Verlust von visuell wahrgenommenen Informationen minimiert werden kann (Benedetto et al. 2011).

Fixationsdauer

Die Bewegung des Blicks stellt ein weiteres Maß für Beanspruchung dar. Für die visuelle Aufnahme von Informationen ist es notwendig, dass ein Mensch die Information fixiert und den Blick stabil auf der Information hält. Eine höhere Dauer dieser Fixationen ist ein Indikator für eine erhöhte mentale Beanspruchung (Bommer und Fendley 2018)

Pupillenweite

Ein weiterer Indikator für Beanspruchung ist die Weite der Pupille. Die Pupillenweite kontrolliert die Menge an Licht, die bei der visuellen Informationsaufnahme durchgelassen wird, weshalb dieses Maß abhängig von den Lichtverhältnissen ist (Beatty und Lucero-Wagoner 2000). Einen zusätzlichen Einfluss auf die Pupillenweite hat die Aufnahme von Koffein (Stuiber 2006). Dennoch stellt ein erhöhter Durchmesser der Pupille einen Indikator für eine erhöhte Beanspruchung dar (Marquart et al. 2015). Ähnlich wie die *Event related Potentials* bei der EEG Messung können mit der ‚*Task-evoked Pupillary Response*‘ Ereignisse in der Aufgabe genutzt werden, um die Pupillenreaktion damit zu korrelieren (Beatty 1982). Zusätzlich konnten (Wang et al. 2013) ein Beanspruchungsmaß auf Basis von Haar-Like Features (Lienhart und Maydt 2002) erproben, das robust gegenüber Beleuchtungsänderungen ist.

Blickverhalten

Eine weitere Möglichkeit, die Vermessung des Auges zu nutzen, um Beanspruchung zu schätzen, ist die Analyse des Blickverhaltens des Menschen und der dabei angesehenen Objekte. Diese Form von Analyse ist allerdings eine kontextbehaftete Messung, die daher in Kapitel 4.3.2 beschrieben wird.

3.2.1.4 Abstraktere Messmethoden

Zusätzlich zu den bereits erwähnten Messmethoden kann die Beanspruchung auch aus abstrakteren Verhaltensparametern ermittelt werden. Hierbei wird die Analyse von Sprache genutzt. Menschliche Ausbilder beschreiben häufig, dass sie die Überbeanspruchung ihrer Schüler bemerken, indem diese nicht mehr sinnvoll auf Sprache reagieren. Diesen Effekt konnten (Berthold und Jameson 1999) auch in Experimenten nachweisen, bei denen die Antwortzeiten der Probanden aufgrund von erhöhter Beanspruchung stark gestiegen sind. Außerdem verändert erhöhte Beanspruchung auch die verwendeten Sprachmuster, indem die Anzahl der syntaktischen Fehler sowie von geäußerten Verzögerungslauten steigt.

Einen weiteren Indikator für Beanspruchung stellen starke Ausdrücke von Emotionen dar (Chaouachi und Frasson 2012). So kann beispielsweise die Erkennung von Verwunderung helfen, Situationen erhöhter Beanspruchung zu identifizieren, da vor allem Experten in der Bedienung eines Systems ausschließlich in beanspruchenden Ausnahmesituationen sich über das Verhalten des Systems wundern. Gleichzeitig nehmen Emotionen einen besonderen Stellenwert im Bereich von adaptiven Systemen ein, da diese trotz ihrer Einordnung in die kontextfreien Ermittlungsmethoden einen besseren Aufschluss darüber geben können, wie eine gegebene Situation zu bewerten ist.

3.2.2 Kontextbehaftete Ermittlungsmethoden

Im Gegensatz zu den kontext-freien Ermittlungsmethoden wird bei den kontext-behafteten Ermittlungsmethoden die Beanspruchungsermittlungsleistung durch Domänenwissen erzeugt. Das Domänenwissen wird dabei genutzt, um kritische Situationen zu definieren oder wiederkehrende Aufgabenmuster zu identifizieren. Durch das Identifizieren von Aufgabenmustern, kann die Art und Weise bewertet werden mit der Mensch seine Aufgaben durchführt und aus diesem Verhalten den Unterstützungsbedarf bestimmen. Außerdem können kritische Situationen ermittelt werden indem analysiert wird, was der Mensch tut. Dies kann verglichen werden mit einem Modell der notwendigen Aufgaben und dadurch besteht die Möglichkeit Fehler in der menschlichen Aufgabendurchführung zu finden.

3.2.2.1 Verhaltensbasierter Ansatz

In einem Experiment konnte (Sperandio 1971) zeigen, dass die gleiche Aufgabe unter hoher Beanspruchung anders durchgeführt wird, als in regulären Beanspruchungssituationen. Die Erkennung dieser, ‚Selbstadaptive Strategie‘ genannten, Verhaltensanpassung in wiederkehrenden Aufgabensituationen kann somit als Beanspruchungsindikator genutzt werden.

Nach (Canham 2019) stellen diese Strategien den Versuch des Operateurs dar, das Erreichen der eigenen Leistungsgrenzen hinauszuzögern. Dabei werden die zwei Varianten ‚*load-shedding*‘ und ‚*load-sharing*‘ unterschieden. Beim ‚*load sharing*‘ übergibt der Mensch einen Teil seiner Aufgaben an einen anderen Menschen oder an eine Automation. Ein Beispiel für diese Art von Strategie kann das Aktivieren des Autopiloten darstellen, wenn der Mensch die Flugaufgabe für gewöhnlich selbstständig durchführt, aber aufgrund der Beanspruchungssituation die Kontrolle übergibt und währenddessen andere Aufgaben durchführt. Der Einsatz dieser Strategie kann dabei unter Inkaufnahme von Leistungseinbußen stattfinden, weil der Mensch die Kontrolle übergibt, obwohl ihm bewusst ist, dass die Automation die Aufgabe schlechter durchführt, als der Mensch es selbst tun würde. Im Gegensatz dazu passt der Mensch beim ‚*load shedding*‘ die Art seiner Aufgabendurchführung an, um die Leistung aufrechtzuerhalten. Eine solche Anpassung konnten (Harris Sr und Mixon 1979) im Blickverhalten von Piloten nachweisen. Die Piloten hatten in einer Flugaufgabe einen stärkeren Fokus auf die Höhenanzeige, wenn die Belastung der Aufgabe erhöht wurde indem ein Kurvenflug statt eines geraden Flugs durchgeführt werden musste. Einen ähnlichen Effekt konnten (Schulte und Donath 2011) für eine Helikoptermission zeigen, in der mehrere unbemannte Flugzeuge kommandiert werden mussten, um die Flugroute des Helikopters aufzuklären und dadurch abzusichern. In dieser Mission wurde neben dem Blickverhalten auch die Nutzung der verfügbaren Werkzeuge angepasst, damit das Erreichen von Überlastungssituationen vermieden werden kann. Die angewandten Strategien der Versuchsteilnehmer beinhalteten eine Reduktion der Aufklärungsleistung, indem ausschließlich eines der drei verfügbaren unbemannten Flugzeuge eingesetzt wurde oder das Zoomlevel der Aufklärungsbilder stark reduziert wurde, wodurch eine Klassifikation der aufgeklärten Objekte nahezu unmöglich gemacht wurde. Außerdem wurde eine Umpriorisierung durchgeführt, bei der die Klassifikation der Bilddaten bewusst verzögert wurde, um andere relevante Aufgaben vorzuziehen. In einer Folgestudie (Donath und Schulte 2015) konnten sie mit Hidden-Markov Modellen eine automatisierte Unterscheidung der Aufgabendurchführung unter Anwendung von selbstadaptiven Strategien zu einer regulären Durchführung der gleichen Aufgabe generieren. Dieser Ansatz erlaubt es einem technischen System hohe Beanspruchungssituationen des Menschen aus seinem Verhalten zu extrahieren und gleichzeitig den relevanten Aufgabenkontext zu ermitteln.

3.2.2.2 Modellbasierter Ansatz

Bei dem modellbasierten Ansatz wird das Wissen über die Aufgabendomäne genutzt, um die Tätigkeit des Menschen zu ermitteln und somit sowohl eine Näherung für den Unterstützungsbedarf zu erzeugen als auch Indikatoren für die Erbringung einer Unterstützungsleistung zu liefern. Ein solches System haben (Maiwald und Schulte 2016)

erfolgreich evaluiert. In ihrem Experiment wurden die Tätigkeit eines Helikopterpiloten auf Basis von Situationsinformationen, wie der Position innerhalb des Missionsszenarios oder dem aktuellen Missionsziel, geschätzt. Die ermittelte Tätigkeit wurde daraufhin mit einem Modell der benötigten mentalen Ressourcen kombiniert, wodurch das technische System die am wenigsten beanspruchten Ressourcenkanäle auswählen konnte, um dort Unterstützungsdialoge zu präsentieren. Diese Anpassung an die freien Beanspruchungsressourcen der Versuchsteilnehmer konnte signifikante Verbesserungen in der Flugleistung gegenüber einer nicht adaptiven Konfiguration erzielen. Des Weiteren konnte mit Hilfe eines NASA-TLX Fragebogens ermittelt werden, dass die Beanspruchung für die Versuchsteilnehmer in der ressourcenangepassten Konfiguration signifikant niedriger war, als in Konfigurationen, in denen die Unterstützungsmeldungen ausschließlich per Text oder ausschließlich per Sprache ausgegeben wurden.

Aufbauend auf diesem Ansatz haben (Brand und Schulte 2021) ein Assistenzsystem entwickelt, das Überlastungssituationen einer Helikoptercrew anhand der Tätigkeit der Piloten ermittelt und daran angepasst Unterstützungen bereit stellt. Das System beruht auf der Tätigkeitsbestimmung von (Honecker und Schulte 2017), die zusätzlich zu den Situationsinformationen des zuvor beschriebenen Ansatzes die Interaktionen der Piloten mit der Mensch-Maschine Schnittstelle für die Ermittlung der Tätigkeit nutzt.

Abbildung 18 visualisiert die Prozesskette der Tätigkeitsbestimmung, deren Kernkomponente das Aufgabenmodell von (Brand und Schulte 2017) bildet. Die Prozesskette beginnt mit der Sammlung der Mensch-Maschine Interaktionen, die daraufhin mit symbolischen Bedeutungen erweitert werden. In Kombination mit dem Aufgabenmodell werden aus diesen Interaktionen Schätzungen über die Aufgabendurchführung der modellierten Aufgaben generiert. Dies erzeugt eine Schätzung über die derzeit ausgeführte Aktivität und mithilfe einer Beanspruchungsmodellierung innerhalb des Aufgabenmodells kann zu dieser Aktivität eine Beanspruchungsschätzung generiert werden.

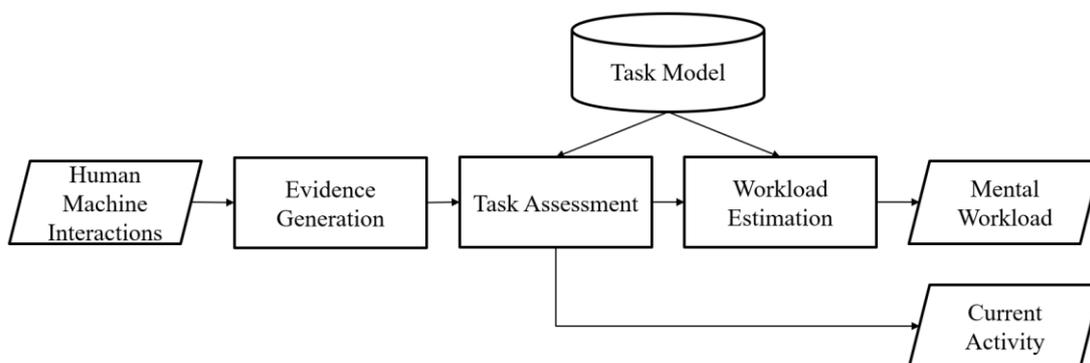


Abbildung 18: Ablauf der modellbasierten Aufgaben und Beanspruchungsanalyse

Das Aufgabenmodell ist eine Repräsentation des Wissens über die möglichen Aufgaben innerhalb eines bestimmten Arbeitsumfelds, sowie den Beziehungen zwischen den Aufgaben. Das Modell beinhaltet hierbei zwei unterschiedliche Arten von Beziehung: Unteraufgaben und Vererbung. Die Unteraufgabenbeziehung beschreibt eine Aufgabe, die die aus mehreren Teilschritten besteht, wie beispielsweise das Abarbeiten einer Checkliste. In diesem Fall wäre jeder Eintrag in der Checkliste eine Aufgabe, die eine Unteraufgabe der Abarbeitung der gesamten Checkliste ist. Die Vererbungsbeziehung ist eher eine strukturierende Beziehung, bei der Aufgaben eines gleichen Typs miteinander verbunden werden, um Gemeinsamkeiten auszunutzen. Hierbei können manuelles Fliegen und das Fliegen mit einem Autopiloten mit einer allgemeineren Flugaufgabe verbunden werden. Dadurch kann leichter modelliert werden, dass beide Aufgaben nur ausführbar sind, wenn das Luftfahrzeug in der Luft ist.

Neben den Beziehungen der Aufgaben zueinander beinhaltet das Aufgabenmodell zusätzliche Informationen über die modellierten Aufgaben. Bei diesen Informationen handelt es sich zum einen um eine Beschreibung der Beanspruchung, die durch die Ausführung der Aufgabe ausgelöst wird. Die Beanspruchung ist modelliert über die acht Beanspruchungsdimensionen aus der, in Kapitel 3.1.4 beschriebenen, Multiple Resource Theory und wird dabei für jede Aufgabe definiert.

Zum anderen werden die Interaktionen beschrieben, die als Indikator für die Aufgabendurchführung dienen. Hierbei werden jeder Aufgabe die relevanten Interaktionen zugeordnet, und ein Zahlentriplett auf Basis der Dempster-Shafer Theorie, einer Erweiterung der Wahrscheinlichkeitstheorie, (Gordon und Shortliffe 1984) definiert. Diese Triplets bestehen aus Werten für Glaube, Zweifel und Ignoranz. Ein hoher Glaubenswert beschreibt dabei, dass die zugehörige Interaktion für die Ausführung einer Aufgabe spricht, während ein hoher Zweifelwert gegen die Ausführung der Aufgabe spricht. Der Ignoranzwert signalisiert eine verbleibende Restunsicherheit, wodurch die drei Werte zu 1 aufsummiert werden. Die Kombination aus Interaktion und dem Dempster-Shafer Triplett (DST) wird als Evidenz bezeichnet und wird genutzt, um eine Schätzung über die Aufgabendurchführung zu erzeugen.

Mithilfe des Aufgabenmodells läuft die Aufgabenerkennung wie folgt ab. Die manuellen und visuellen Interaktionen werden auf Basis der zugewiesenen DST in Evidenzen umgewandelt und bilden somit Indikatoren für die Durchführung der einzelnen Aufgaben. Diese Evidenzen werden daraufhin mit einer vereinfachten Form der Dempster-Shafer Kombinationslogik zu einer Schätzung für die jeweiligen Aufgaben kombiniert. Die vereinfachte Form besteht hierbei daraus, dass die Evidenzen paarweise nacheinander kombiniert werden, anstatt alle Evidenzen in einem Schritt zu kombinieren. Sie ist

notwendig, damit die Berechnung in Echtzeit durchgeführt werden kann. Das Ergebnis der Kombination der Evidenzen ist erneut ein DST für jede mögliche Aufgabe, welches die Schätzung zu der Durchführung der jeweiligen Aufgabe beschreibt. Da eine solche Schätzung in jedem Zeitschritt berechnet werden kann, kann zu jeder Aufgabe der Verlauf dieser Schätzungen angezeigt werden, wie es Abbildung 19 schematisch dargestellt ist.

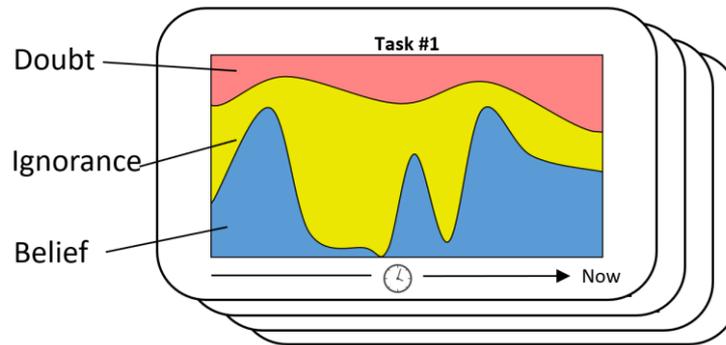


Abbildung 19: Schematische Darstellung der Aufgabenhypothesentriplets

Zu jeder Aufgabe, deren Glaubenswert über einem Schwellwert von 0,5 liegt, wird daraufhin angenommen, dass sie zu dem zugehörigen Zeitpunkt durchgeführt wird. Die Sammlung aller aktiven Aufgaben wird als Aktivität bezeichnet. Die Bestimmung der Aktivität erlaubt es zusätzlich eine Schätzung empfundenen Beanspruchung des Probanden zu erzeugen. Zur Beanspruchungsermittlung wurden hierbei die zwei Methoden von (Maiwald 2015) genutzt, um sowohl eine Schätzung für die einzelnen Beanspruchungsdimensionen als auch eine Gesamtbeanspruchungsschätzung zu generieren. Für jede der einzelnen Beanspruchungsdimensionen wird dabei die Summe dieser Dimension über alle aktiven Aufgaben gebildet, während die Gesamtbeanspruchung über den W/Index aus Kapitel 3.1.4 beschrieben wird. Abbildung 20 stellt das Resultat einer solchen Schätzung dar. Oben in der Abbildung ist zu sehen, dass die durchgeführten Aufgaben hierbei vorwiegend die Dimensionen: ‚visuell-räumliche‘ Informationsaufnahme, die ‚kognitiv-räumliche‘ Informationsverarbeitung und die ‚manuelle‘ Reaktion beanspruchen. Im unteren Teil der Abbildung sind die, bei der Aufgabendurchführung auftretenden, Ressourcenkonflikte, sowie die resultierende Gesamtbeanspruchung zu sehen.



Abbildung 20: Beispielhafte Beanspruchungsbeschreibung als Ergebnis der Aufgabenanalyse

3.3 Konsolidierung

Die Operationalisierung der menschlichen Beanspruchung basiert auf Modellen von dem Zusammenhang zwischen Beanspruchung und der Aufgabendurchführung oder den Reaktionen des menschlichen Körpers. In der Literatur existiert eine Vielzahl von Körperreaktionen auf mentale Beanspruchung. Insbesondere die Herzrate und die augenbasierten Beanspruchungsindikatoren bilden ein robustes indirektes Maß. Andererseits besteht bei diesen kontextfreien Methoden häufig das Problem, dass eine Änderung des jeweiligen Parameters auch durch unabhängige Einflüsse ausgelöst werden kann, wie beispielsweise die Lichtverhältnisse bei den augenbasierten Messungen. Daher existieren bisher nur wenige Veröffentlichungen bei denen eine Beanspruchungssituation, gemessen durch kontextfreie Ermittlungsmethoden, zu Echtzeitanpassungen des technischen Systems führt. Nichtsdestotrotz zeigen das Framework RASMUS (Schwarz und Fuchs 2018) und die Versuche von (Dorneich et al. 2012), dass diese Formen der adaptiven Automation erfolgreich eingesetzt werden können.

Auf der Seite der kontextbehafteten Messungen zeigen sich ebenfalls vielversprechende Möglichkeiten die Automation an die Bedürfnisse der menschlichen Operateure anzupassen und somit die Gesamtleistung des Mensch-Maschine Systems zu erhöhen. Die Schwierigkeit dieser Methoden ist der notwendige Modellierungsaufwand, für die Beschreibung der Aufgabendomäne, und dass die verhaltensbasierten Methoden Aufgabensituationen benötigen, die sich häufig wiederholen.

Für den Einsatz in der MUM-T Zielanwendung stellt sich daher die Frage, welche der vorgestellten Methoden sich eignet. Daher werden im Folgenden mehrere Experimente durchgeführt, die die Bewertung der einzelnen Operationalisierungsmethoden zum Ziel haben. Der Fokus der Bewertung liegt dabei auf der Eignung für den Einsatz in einem closed-loop adaptiven Automationssystem.

4. Untersuchung psychophysiologischer Messmethoden

Dieses Kapitel beinhaltet die Experimente, die im Rahmen dieser Arbeit durchgeführt wurden, um die Nutzbarkeit der einzelnen Methoden für adaptive Automation bewerten zu können. Dazu wird im Folgenden zuerst der Kampffjetsimulator des Instituts für Flugsysteme sowie die dazugehörige Zielsetzung beschrieben, da dieser Simulator als Experimentalumgebung dient. Infolgedessen werden die gewählten Bewertungskriterien vorgestellt aufgrund derer diskutiert werden soll, in welchem Fall welche Form von Messung den Einsatz adaptiver Automation unterstützen kann. Abschließend werden die unterschiedlichen Arten von psychophysiologischen Messmethoden vorgestellt, die dazu durchgeführten Experimente beschrieben und eine Bewertung der jeweiligen Messmethode abgegeben.

4.1 Bewertungskriterien

Der Rest dieses Kapitels beschäftigt sich mit der Kategorisierung, Beschreibung und Bewertung unterschiedlicher psychophysiologischer Messmethoden hinsichtlich ihres Nutzens für adaptive Automation. Um besagte Bewertung durchführen zu können werden im Folgenden die zugrundeliegenden Bewertungskriterien beschrieben und voneinander abgegrenzt. Die ersten drei Kriterien sind hierbei gängige Bewertungskriterien für die Messung von mentaler Beanspruchung (Heard et al. 2018), während die letzten drei Kriterien im Rahmen dieser Arbeit hinzugefügt wurden um eine Eignung der Messmethode für den Einsatz in adaptiven Automationssystemen zu bewerten.

4.1.1 Robustheit/Selektivität

Das Robustheitskriterium beschreibt die Anfälligkeit der Messmethode auf etwaige Störeinflüsse. Bei einer sehr robusten Messmethode kann eine Aussage über die Beanspruchung sehr zuverlässig abgeleitet werden, weil Störungen wie Rauschen die Aussage der Methode nicht beeinflussen oder herausgerechnet werden können. Einen weiteren Aspekt der Robustheit (in der Literatur als Selektivität bezeichnet) bildet die Abgrenzbarkeit von unabhängigen Belastungsfaktoren wie emotionalem Stress. Methoden mit einer geringen Robustheit können nur in kontrollierten Laborumgebungen eingesetzt werden. So ist beispielsweise die Messung der Pupillenweite zur Beanspruchungsbestimmung eine sehr empfindliche Messung, da die Pupillenweite neben der Beanspruchung auch stark von den Lichtverhältnissen beeinflusst wird.

4.1.2 Sensitivität

Dieses Kriterium beschreibt die mögliche Unterscheidungstiefe der Beanspruchungsmessmethode. Hohe Sensitivität zeichnet sich hierbei dadurch aus, dass viele Abstufungen in der Beanspruchungsklassifikation durchgeführt werden können. Durch die Vielseitigkeit der Zusammenhänge zwischen den Reaktionen des Körpers und der Beanspruchung ist die Sensitivität von psychophysiologischen Messmethoden hierbei grundsätzlich niedrig einzustufen im Vergleich zu Fragebogenmethoden oder statischen Modellierungen wie beispielsweise bei dem VACP-Modell von Aldrich und McCracken in Kapitel 3.1.7. Diese Arten der Beanspruchungsschätzung eignen sich allerdings nicht für den Einsatz in adaptiver Automation.

4.1.3 Diagnostizität

Mit dem Diagnostizitätskriterium wird beschrieben, ob eine Art von Beanspruchung genauer festgestellt werden kann oder lediglich ein Grundverständnis von Beanspruchung bewertet wird. Wie bereits in Kapitel 3.1.2 beschrieben werden in den Beanspruchungsmodellen mehrere Arten von Beanspruchung unterschieden. Die häufig differenzierten Arten von Beanspruchung sind *visuell*, *auditiv*, *kognitiv* und *psychomotorisch*. Eine Messmethode mit einer hohen Selektivität bedeutet hierbei, dass eine bestimmte Art von Beanspruchung genauer festgestellt werden kann. Ein Beispiel hierfür stellt die Task Evoked Pupillary Response dar (Beatty 1982), bei der Änderungen der Pupillenweite festgestellt werden, sobald die Informationen für eine neue Aufgabe wahrgenommen werden.

4.1.4 Individualität

Mit der Individualität der Messmethode wird bewertet, wie stark auf individuelle Unterschiede zwischen den unterschiedlichen Probanden eingegangen werden kann. Hierbei ist vor allem wichtig, dass in einer kritischen Situation ermittelt werden kann, dass der betroffene Versuchsteilnehmer überbeansprucht ist, anstatt aus der Situation eine Beanspruchung zu schließen. Eine hohe Individualität entspricht einem sehr genauen Abbild des individuellen Probanden, gleichzeitig bedeutet sie aber auch einen hohen Kalibrierungsaufwand, um die individuellen Unterschiede nutzbar zu machen.

4.1.5 Generalisierbarkeit

Dieses Kriterium beschreibt, wie übertragbar Erkenntnisse aus einem Anwendungsfall oder Experiment sind. Grundsätzlich gilt hierbei: Je mehr Domänenwissen notwendig ist um eine Methode zur Beanspruchungsschätzung durchzuführen, desto schlechter lassen sich die

Ergebnisse übertragen und somit desto schlechter generalisierbar ist die Methode. Physiologische Zusammenhänge zu Beanspruchung, wie beispielsweise eine steigende Herzrate bei hoher kognitiver Beanspruchung, eignen sich daher als Maße in Anwendungsfällen bei denen die Generalisierbarkeit wichtig ist.

4.1.6 Direkte Anwendbarkeit

Das letzte Kriterium beschreibt, ob die Analyse des Beanspruchungsmaßes gleichzeitig offenlegt, wie eine adaptive Automation geartet sein sollte um den Operateur effektiv zu unterstützen. Laut (Fuchs et al. 2006) ist hierzu zusätzliche Kontextinformation notwendig bei der das zugrunde liegende Problem des Operateurs im System repräsentiert werden kann. Das Wissen, dass der Operateur in einer Überbeanspruchungssituation ist, ist nutzlos, wenn keine Maßnahmen getroffen werden können, um die Beanspruchung zu senken.

4.2 Manned-unmanned Teaming Kampffjetsimulator

Als Experimentalplattform dieser Arbeit dient der, in Abbildung 21 dargestellte, *Manned-unmanned Teaming* Kampfflugzeugsimulator des Instituts für Flugsysteme (IFS) der Universität der Bundeswehr München. Die Zielsetzung für diesen Simulator ist es, die Kooperation eines menschlichen Piloten an Bord eines Kampfflugzeugs mit mehreren unbemannten Kampfflugzeugen zu untersuchen. In dieser Untersuchung soll ermittelt werden, ob der Pilot durch ausreichende Unterstützungsfunktionen in Form von adaptiver Automation im bemannten Cockpit, sowie kognitiver Automation an Bord der unbemannten Flugzeuge, dazu ermöglicht werden kann dieses bemannt-unbemannte Team allein in militärischen Einsatzszenarien befehligen zu können. Gleichzeitig werden hierbei ebenfalls Methoden untersucht die notwendige adaptive Automation zu verbessern.



Abbildung 21: MUM-T Kampffjetsimulator des Instituts für Flugsysteme

Die Simulationsumgebung besteht aus einem Cockpit mit einem kampffjetüblichen „Hands on throttle-and-stick“ System der Firma Reiser (Reiser Simulation and Training GmbH) und einem Dome-Projektionssichtsystem. Für die Führung der unbemannten Kampfflugzeuge und die Durchführung der Missionsmanagement Aufgaben hat das Cockpit zusätzlich drei Touch-Bildschirme mit jeweils einer C++ Software mit einer grafischen Benutzerschnittstelle. Die Software wurde in der Zeit dieser Arbeit am Institut für Flugsysteme entwickelt und simuliert MUM-T Kampfmissionen für die Untersuchung der Führung unbemannter Teammitglieder aus einem Cockpit heraus. Die Zielsetzung bei der Entwicklung war eine möglichst realistische Abbildung echter Einsatzszenarien, damit der Einsatz von adaptiven Automationssystemen in einem closed-loop evaluiert werden kann.

4.2.1 Missionsszenario

Die Simulation bildet „Offensive Counter Air“ Missionen ab, bei denen taktische Bodenziele im Feindgebiet bekämpft werden müssen. Eine Beispielmision ist die Zerstörung einer feindlichen Flugzeugstartbahn, damit die dort stationierten Flugzeuge nicht starten können und dadurch eine Luftüberlegenheit der eigenen Kräfte hergestellt werden kann. Die Komposition der eigenen Kräfte besteht dabei aus einem bemannten Flugzeug und für gewöhnlich drei unbemannten Teammitgliedern. Das Missionsgebiet wird von der „Forward line of own troops“ (FLOT) zweigeteilt. Auf der einen Seite der FLOT befinden sich die befreundeten Kräfte, während auf der anderen Seite das Hoheitsgebiet des simulierten Feindes befindet. Im Feindgebiet befindet sich eine Vielzahl von bodengestützter Luftabwehr, die im Rahmen der Mission unterdrückt oder aber zerstört werden müssen, damit die eigenen Flugzeuge bis in das Zielgebiet vordringen können und dort die geplante

Wirkung zu erzielen. Als zusätzliche Randbedingung können sich feindliche Flugzeuge im Feindgebiet befinden, die zusätzlich bekämpft werden müssen, um ein sicheres Vordringen in das Zielgebiet zu garantieren.

4.2.2 Mensch-Maschine Schnittstelle

Für die Planung und Durchführung der Mission besitzt die Mensch-Maschine Schnittstelle dabei eine taktische Kartendarstellung als zentrales Interaktionselement. Die Kartendarstellung wurde mit Hilfe der Kartenengine ‚*Carmenta*‘ (*Carmenta*) entwickelt und stellt die taktisch relevanten Elemente in einer Draufsicht in Echtzeit dar. Die Darstellung der visualisierten Elemente besteht dabei aus den Symbolen des Militärstandards MIL-STD-2525C, um die notwendige Trainingszeit der Versuchsteilnehmer zu minimieren.

Für zusätzliche Informationen kann die Karte mit mehreren ‚*Pages*‘ überlagert werden, die die weiteren Parameter der eigenen Kräfte darstellen und den Zugriff auf die einzelnen Subsysteme des bemannten Flugzeugs erlauben. Die Pages des zentralen Displays sind:

Primary Flight Display: Das ‚*Primary Flight Display*‘ (PFD) stellt den aktuellen Flugzustand des eigenen Flugzeugs dar. Die dargestellten Informationen sind die aktuelle Fluglage (Rotation im dreidimensionalen Raum), die Fluggeschwindigkeit und die Flughöhe, sowohl als Höhe über Normalnull als auch als Höhe über dem Grund.

Defensive Aids Subsystem (DASS): Die DASS-Page erlaubt die Kontrolle über den Einsatz von ‚*Chaffs*‘ und ‚*Flares*‘, die als Täuschkörper dienen, um potentielle feindliche Luft-zu-Luft Raketen abzulenken.

Waffensystem: Die Page für das Waffensystem zeigt an, wie viele Waffen an Bord des eigenen Flugzeugs zur Verfügung stehen und wo die jeweiligen Waffen am eigenen Flugzeug befestigt sind. In der Simulation werden dabei drei Waffentypen unterschieden: ‚*High-speed Anti Radiation Missiles*‘ (HARM) für die Unterdrückung bodengestützter Luftabwehr, Luft-zu-Luft Raketen für die Bekämpfung feindlicher Flugzeuge und ‚*Guided Bombs*‘ für die Zerstörung von Bodenzielen.

Autopilot: Die Autopilot-Page bietet die Möglichkeit eine Zielhöhe, Zielrichtung und Zielgeschwindigkeit an den Autopiloten des bemannten Flugzeugs zu übergeben und den Autopiloten zu aktivieren.

Fuel: Eine Anzeige des verfügbaren Sprits, sowie dessen Verbrauch.

Fighter-System: Die Fighter-System-Page bietet die Kontrolle über Basissysteme des Flugzeugs, wie die Parkbremse und das Fahrwerk.

Flight-Map: Eine weitere Kartendarstellung, die auf das bemannte Flugzeug zentriert ist, damit der Pilot das eigene Flugzeug überwachen kann, während er beispielsweise einen Planungsvorgang auf der Hauptkartendarstellung durchführt.

Target Identification: Die Target Identification-Page besteht aus zwei Teilen. Zum einen bietet sie die Möglichkeit einen Livestream des Kamerabilds von einem der unbemannten Flugzeuge darzustellen, zum anderen präsentiert es die taktischen Elemente der Simulation in einer Listendarstellung, worüber Aufklärungsbilder der jeweiligen Objekte geöffnet werden können, sofern diese existieren.

Timeline: Die Timeline-Page visualisiert den Plan der eigenen Kräfte. Zusätzlich kann sie dazu genutzt werden, in Kombination mit der taktischen Karte, neue Aufträge an die unbemannten Teammitglieder zu erteilen. Ein automatisierter Planer, der im Rahmen der Promotionsverfahrens von Felix Heilemann (Heilemann und Schulte 2020) entwickelt wurde, unterstützt bei der Zuweisung der Aufträge. Nach der Spezifikation eines Auftrags in der taktischen Karte, bietet die Timeline-Page drei Möglichkeiten: Eine Zuweisung an ein unbemanntes Flugzeug zu einem fixen Zeitpunkt durch die Auswahl des Zeitpunkts; Die Zuweisung an ein Teammitglied ohne Vorgabe eines Zeitpunkts; Eine Zuweisung an das Team ohne Vorgabe welches Teammitglied den Auftrag durchführt. Für fehlende Vorgaben berechnet der automatisierte Planer eine optimale Lösung.

Zusätzlich zu dem zentralen Bildschirm (MHDD) befinden sich zwei weitere Bildschirme (SHDD) in dem Kampfflugzeugcockpit, die jeweils rechts und links neben dem zentralen angebracht sind. Auf diesen Bildschirmen befinden sich weitere Pages, von denen pro Bildschirm immer jeweils eine aktiv sein kann:

Communication: Die Communication-Page bietet die Auswahl zwischen mehreren Funkkanälen für die Kommunikation mit anderen Teilnehmern, wie beispielsweise einem Operateur Arbeitsplatz, der einen Tower simuliert. Zusätzlich beinhaltet die Page Informationen über den aktuellen Auftrag.

Engine and Fuel: Diese Page bildet eine Redundanz für die Fuel-Page des MHDD, damit die Informationen betrachtet werden können, ohne die Karte zu verdecken. Weiterhin stellt sie Drehzahl und Öldruck Informationen dar.

Route: Die Route-Page bietet eine weitere Kartendarstellung für eine genauere Visualisierung der geplanten Routen der einzelnen unbemannten Teammitglieder.

Autopilot und PFD: Diese beiden Pages bilden erneut die Möglichkeit zur redundanten Informationsdarstellung und haben die gleichen Interaktionsmöglichkeiten wie ihre Variante auf dem MHDD.

4.2.3 Simulatorarchitektur

Für die Bereitstellung der Simulation müssen mehrere Simulationsaufgaben durchgeführt werden. Aufgrund der benötigten Rechenlast müssen diese Aufgaben außerdem auf mehrere Computer verteilt werden. Dabei können einige Aufgaben, wie beispielsweise die Missionsplanung, auf beliebigen Computern laufen, während andere Aufgaben, wie die Darstellung der Benutzerschnittstelle, an spezifische Computer gebunden sind. Im Folgenden werden die Hardware- und die Softwarearchitektur der Simulation genauer betrachtet.

4.2.3.1 Hardwarearchitektur

Der Kern der Hardwarearchitektur besteht aus einer Garantie einer sicheren Kommunikation, sowie verfügbarem Speicherplatz, durch einen Gigabit-Switch und ein daran angeschlossenes „*Network attached storage*“ (NAS) – System. Damit verbunden sind mehrere Computer mit speziellen Funktionalitäten. Drei dieser Computer sind an jeweils einen Beamer eines Dome-Aussensichtsystems angeschlossen und sind ausschließlich für die Anzeige der Außensicht zuständig. Zwei zusätzliche Computer sind an zwei Fernseher angeschlossen, die eine Beobachtung des Simulationsgeschehens durch die Versuchsleiter ermöglichen, indem sie unter anderem die relevanten Anzeigen der Simulation spiegeln. Ein weiterer Computer ist für die Mensch-Maschine Schnittstelle zuständig und ist sowohl an die Touch-Bildschirme als auch an das HOTAS-System, sowie die Pedale angeschlossen. Der letzte spezialisierte Computer bietet die Schnittstelle zu den Blickbewegungskameras (siehe Kapitel 4.2.4). Wie in Abbildung 22 dargestellt sind die beschriebenen- sowie mehrere zusätzliche Computer über einen KVM-Switch mit einigen Arbeitsplätzen für die Entwicklung der Simulation verbunden, wodurch jeder Arbeitsplatz Zugriff auf jeden der Computer hat.

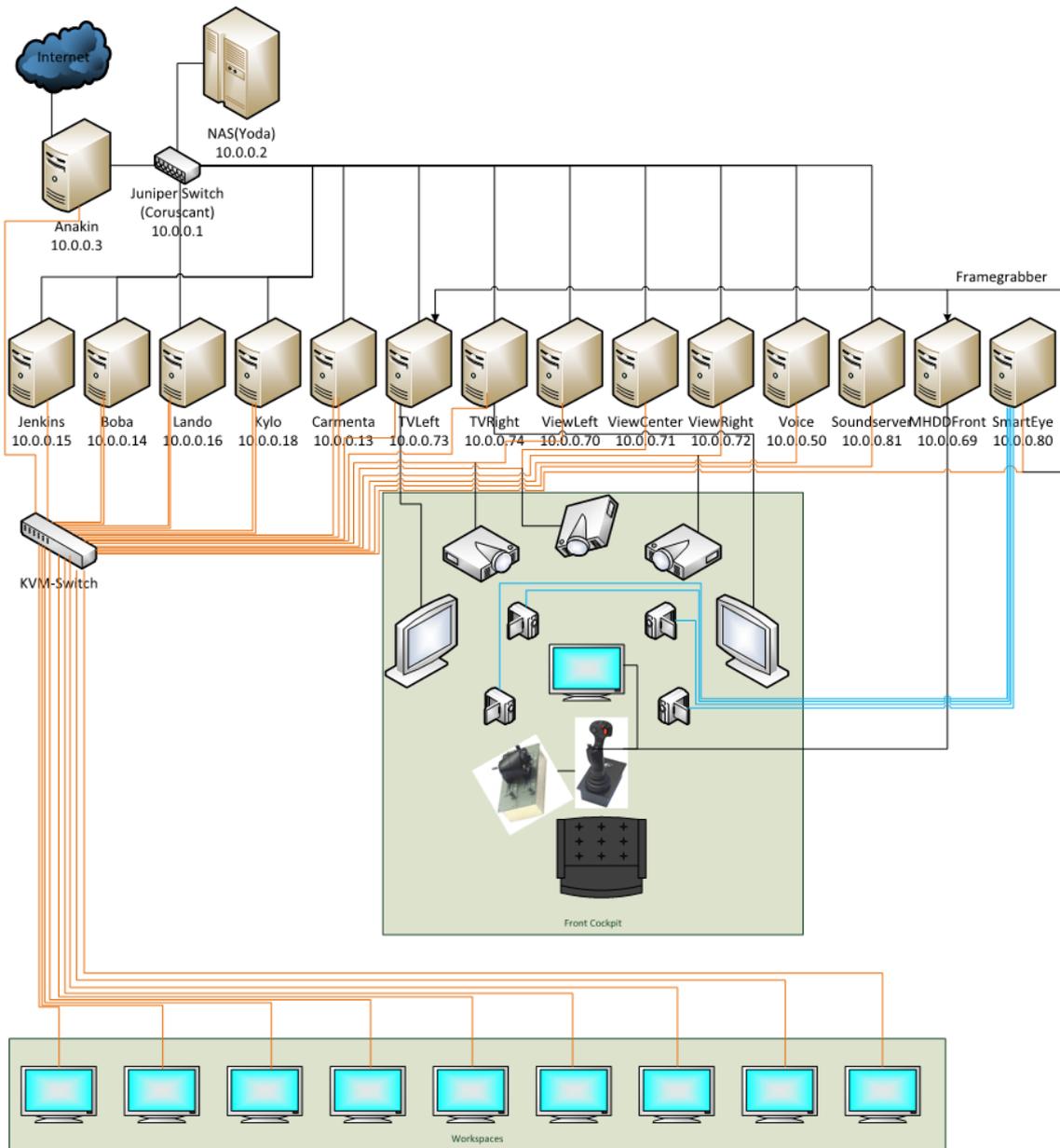


Abbildung 22: Hardwarearchitektur des Kampffjetsimulators

4.2.3.2 Softwarearchitektur

Aufgrund der Notwendigkeit der Verteilung der einzelnen Programme steht bei der Softwarearchitektur die Kommunikation zwischen den Programmen im Vordergrund. Ein Großteil der notwendigen Informationen, wie beispielsweise die Simulationszeit, wird von nahezu allen Programmen benötigt. Daher werden die Informationen in einem gekapselten Modul, mit dem Namen ‚*Model*‘, verwaltet und dieses Modul hält sich selbstständig auf einem aktuellen Stand. Dazu kommuniziert das Model mit einem weiteren Modul für die Interprozesskommunikation (IPC), welches eine Trennung zwischen simulationsspezifischen Datentypen und der eigentlichen Netzwerkkommunikation darstellt.

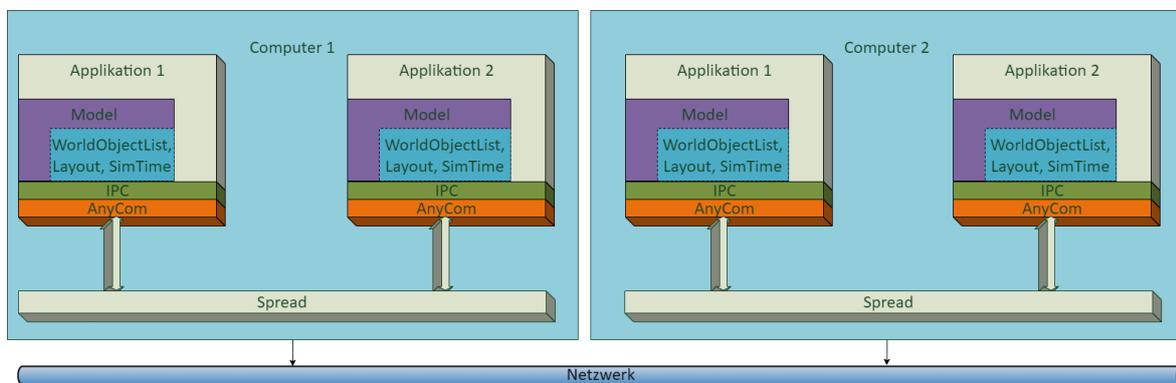


Abbildung 23: Kommunikationsschema der Programme des Kampfflugzeugsimulators

Abbildung 23 stellt dieses Kommunikationsschema graphisch dar. Als Kommunikationsschnittstelle wird eine Eigenentwicklung des Instituts für Flugsysteme genutzt („AnyCom“), die es erlaubt beliebige Nachrichten zwischen Programmen sowohl auf demselben Computer als auch auf unterschiedlichen Computern zu übertragen. Zur Verteilung greift AnyCom dabei auf das Spread-Toolkit (Spread Concepts LLC) zurück, dessen Token-Ring Verfahren dafür sorgt, dass jedes Programm sich mit einem Daemon auf dem eigenen Computer verbindet und dieser die Nachrichten an alle Interessenten verteilt.

Die synchronisierten Daten bestehen vorwiegend aus den drei Komponenten ‚*WorldObjectList*‘, ‚*Plan*‘ und ‚*Layout*‘. Die *WorldObjectList* ist eine Beschreibung der taktischen Lage, die für jedes Objekt die simulierten Informationen enthält. Die Liste ist selbst ebenfalls dreigeteilt in die ‚*GroundTruthList*‘, die die exakten Informationen der Objekte enthält, und eine *WorldObjectList* für die eigenen Kräfte („*BlueWOList*“) und eine Liste für die Feindkräfte („*RedWOList*“), die den Wissensstand der jeweiligen Seite repräsentieren. Die Plan-Datenstruktur enthält die Aufträge der bemannten und unbemannten Flugobjekte und teilt sich in den aktiven Plan, der von den Flugzeugen ausgeführt wird, und den ‚*ModifyPlan*‘, der von den Piloten parallel zu der Auftragsdurchführung editiert werden kann. Das Layout beinhaltet Informationen über die

Konfiguration der Anzeige, insbesondere die derzeit aktiven Pages und die Einstellungen der taktischen Kartendarstellung wie das Zentrum der Karte und die Zoomstufe.

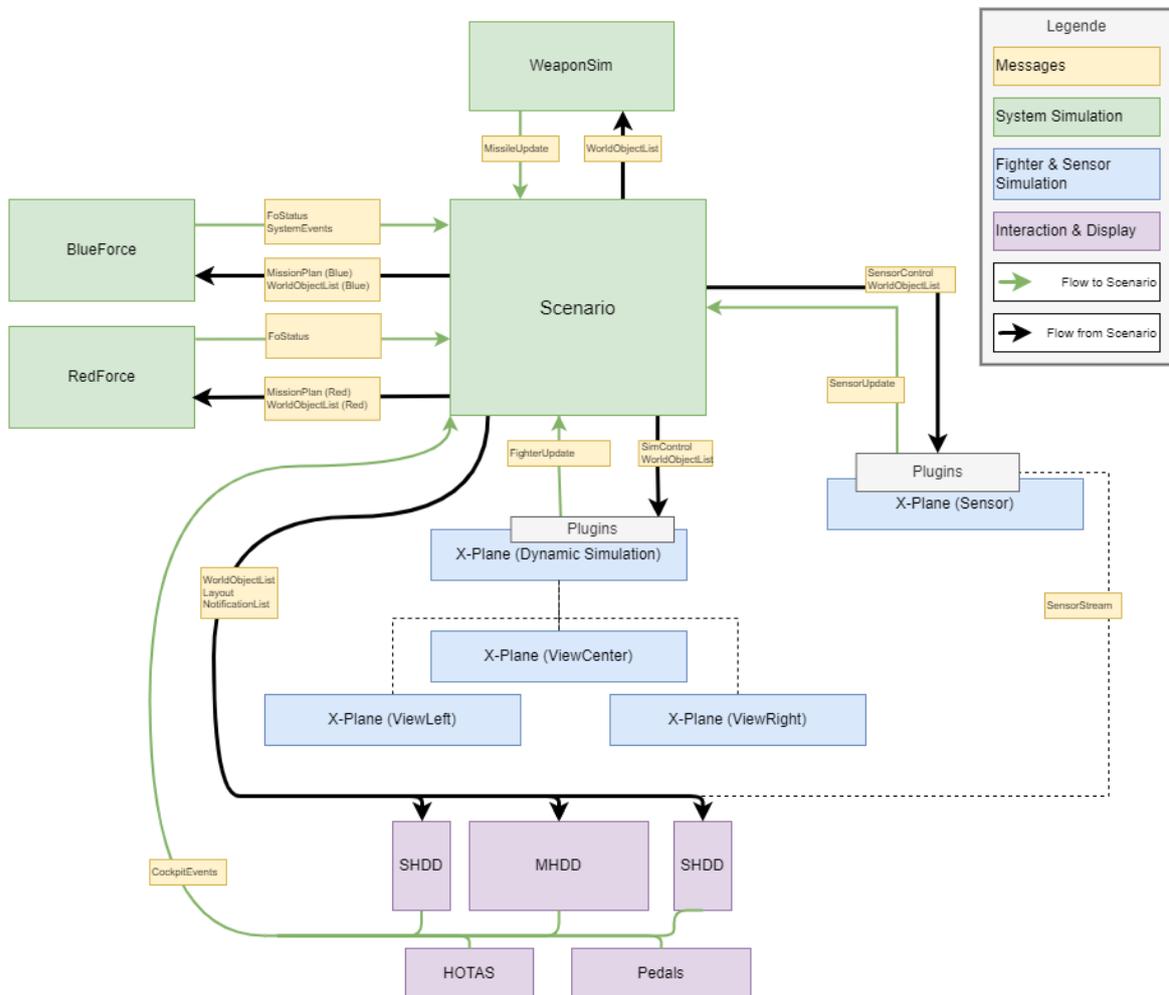


Abbildung 24: Softwaremodularchitektur des Kampfflugzeugsimulators

Die Verwaltung der beschriebenen Daten findet in dem Programm ‚*Scenario*‘ statt. Dieses Programm vereinigt die Informationen der anderen Programme und kreiert daraus die GroundTruthList. Außerdem beinhaltet es das Modul ‚*Obfuscator*‘, das die Sensorsimulation der Weltobjekte durchführt und dadurch die RedWOList und die BlueWOList generiert. Diese Listen werden daraufhin an alle anderen Programme verteilt, die darauf aufbauend den nächsten Zeitschritt berechnen können. Die GroundTruthList wird an alle Programme verteilt, die die physikalische Simulation durchführen. Die wichtigsten dieser Applikationen sind die ‚*WeaponSim*‘ für die Simulation der Raketen, eine Simulation der Flugdynamik des bemannten Flugzeugs mit Hilfe der Software X-Plane 11 (Laminar Research) und die Simulation der unbemannten Einheiten. An dieser Stelle bedeutet unbemannte Einheiten sowohl die Simulation der feindlichen Kräfte als ‚*Computer*

Generated Forces in der Applikation *RedForce*, als auch die Simulation der unbemannten Teammitglieder in der Applikation *BlueForce*. Diese beiden Applikationen simulieren zum einen die Physik der zugehörigen Einheiten, zum anderen aber ebenfalls deren Verhalten. Die Verhaltensgenerierung wurde dabei von Sebastian Lindner (Lindner und Schulte 2020) als eines der Kernthemen seines Promotionsvorhabens entwickelt. Ein wichtiges Element der Verhaltensgenerierung ist die Interpretation der taktischen Lage, weshalb die BlueForce sowie die RedForce zusätzlich zu der GroundTruthList die jeweilige Liste der simulierten Kräfte benötigt. Neben der Verhaltensgenerierung der unbemannten Einheiten wird die BlueWOList außerdem für alle Anzeigen der Mensch-Maschine Schnittstelle genutzt, damit garantiert werden kann, dass dem Piloten keine Informationen dargestellt werden, die den eigenen Kräften nicht bekannt sind.

4.2.4 Blickbewegungsmesssystem

Die Simulationsumgebung ist ausgestattet mit einem Blickbewegungsmesssystem der Firma SmartEye (Smart Eye AB), wie in Abbildung 25 dargestellt. Hierbei handelt es sich um ein kamerabasiertes Messsystem bestehend aus vier Infrarotkameras sowie drei Infrarotstrahlern. Für die Blickbewegungserfassung wird hierbei ein video-basiertes Blickbewegungsmessverfahren auf Basis der Cornea-Reflex Methode eingesetzt (Di Gao et al. 2012). Bei diesem Verfahren wird Infrarotlicht in das menschliche Auge gestrahlt und die Reflektion dieses Lichts auf der Hornhaut mit einer Kamera aufgenommen. Dies erlaubt die Bestimmung der Blickrichtung durch den Abstand dieser Reflektion zu der Pupille des Menschen.



Abbildung 25: Blickbewegungsmesssystem im Kampffjetsimulator

Die Zuordnung der Blickrichtung zu einer Position im Simulator Cockpit entsteht hierbei durch eine Modellierung des Cockpits in Relation zu den Kameras und den Infrarotstrahlern. Dieses, als Weltmodell bezeichnete, Modell ist in Abbildung 26 dargestellt und erlaubt eine

Berechnung des Durchstoßpunktes der Blickrichtung mit den vorhandenen Anzeigeelementen.

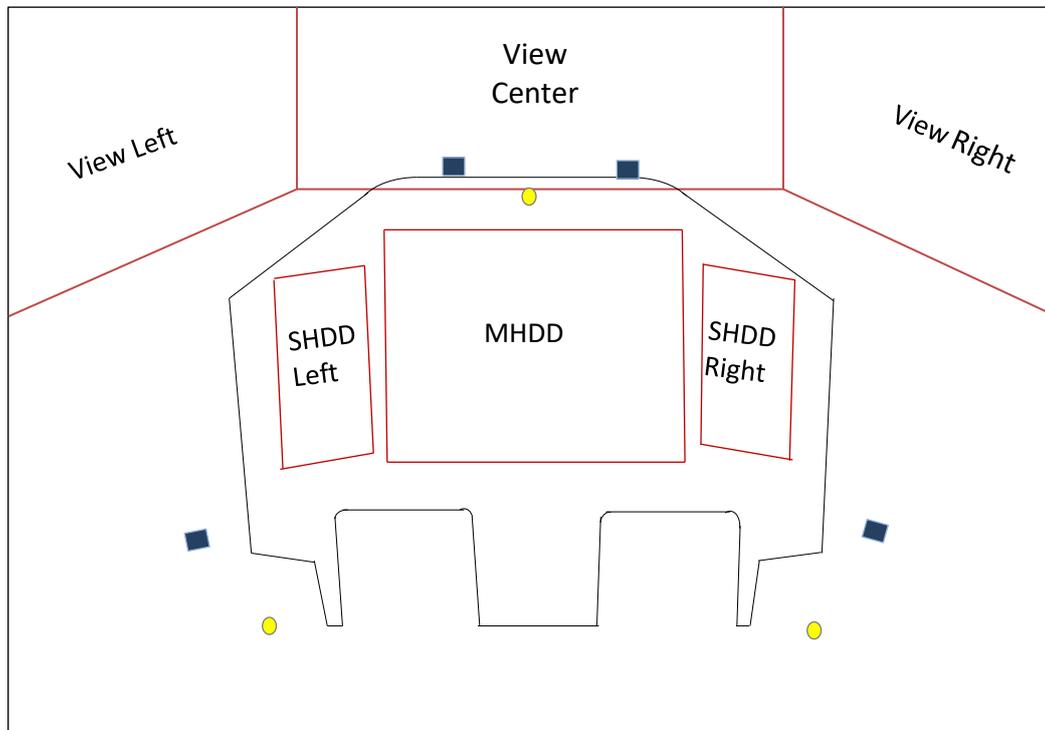


Abbildung 26: Weltmodell der Blickbewegungsmessung

4.2.4.1 Positionierung der Kameras

Für die Ermittlung der Blickrichtung musste eine optimale Position sowohl für die vier Infrarotkameras des Blickbewegungsmesssystems als auch für dessen Infrarotstrahler gefunden werden. Für die Positionierung muss beachtet werden, dass die Strahler einen Abstand von mindestens 15cm zu jeder Kamera haben muss, um zu verhindern, dass eine ‚Bright Pupil‘ erzeugt wird. Bei diesem Effekt trifft das Licht der Infrarotstrahler in einem speziellen Winkelverhältnis auf die Pupille, sodass sie wie in Abbildung 27 im Kamerabild weiß erscheint. Zusätzlich müssen die Kameras ein möglichst großes Sichtfeld abdecken, während gleichzeitig garantiert ist, dass jederzeit mehrere Kameras die Augen erfassen können. Zu diesen Rahmenbedingungen kommt erschwerend hinzu, dass durch die Bedienung der Touch-Bildschirme häufig die Hände des Menschen, die Kameras verdecken. Außerdem sollten die Kameras im unteren Bereich des Sichtfelds platziert werden, weil das Augenlid einen größeren Teil des Auges verdeckt, wenn die Kameras von oben auf das Auge gerichtet sind.

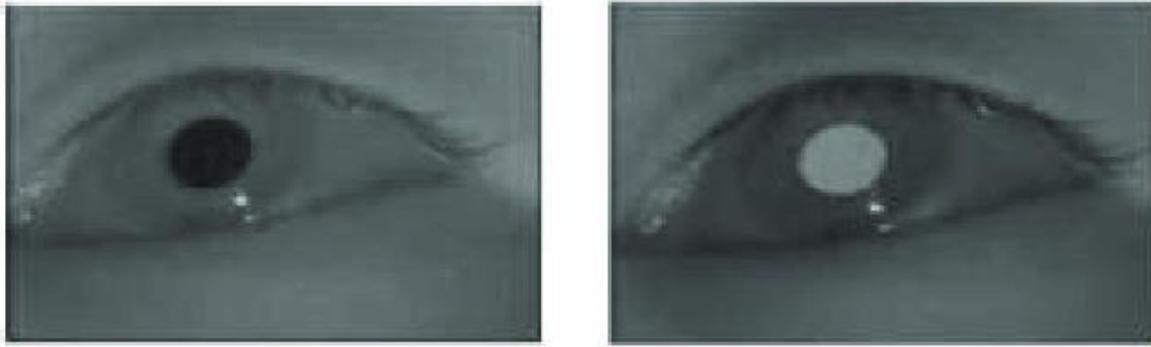


Abbildung 27: Normale Pupille (links) im Vergleich zu einer 'Bright Pupil' (entnommen aus (Al-Rahayfeh und Faezipour 2013))

4.2.4.2 Erstellung eines Weltmodells

Nachdem die Kameras und die Infrarotstrahler positioniert sind müssen der Software des Blickbewegungsmesssystems die Positionen mitgeteilt werden. Dazu müssen zuerst die Kamerapositionen zueinander kalibriert werden. Dabei wird ein Schachbrettmuster mit bekannter Größe vor den Kameras positioniert. Dieses Schachbrett kann in den jeweiligen Kamerabildern erkannt werden und durch die bekannte Größe der einzelnen Schachfelder können die Eckpunkte des Schachbretts in als dreidimensionale Koordinaten in den jeweiligen Kamerakoordinatensystemen berechnet werden. Dieser Vorgang wird in allen Kamerabildern gleichzeitig durchgeführt und somit können die relativen Posen der Kameras zueinander bestimmt werden. Aufbauend auf dieser extrinsischen Kamerakalibrierung wird ein Koordinatenursprung für das Weltkoordinatensystem bestimmt. Dazu wird erneut das Schachbrett vor den Kameras positioniert, wobei dieses Mal die Position des Schachbretts reproduzierbar sein muss (z.B. durch deine mechanische Fixierung). Das Zentrum des Schachbretts wird als neuer Koordinatenursprung des Weltmodells festgesetzt.

Neben dem Koordinatenursprung und den Kameras beinhaltet das Weltmodell zusätzlich die Positionen der Infrarotstrahler und die Bildschirme. Für die Bestimmung dieser Positionen wird erneut ein Schachbrett benutzt. An diesem Schachbrett ist ein Laserentfernungsmesser befestigt, wodurch die Entfernungsmessung des Messgeräts auf die Pose des Schachbretts aufaddiert werden kann, um einen 3D-Punkt zu lokalisieren der nicht von den Kameras gesehen werden kann. Mit dieser Methode können die Eckpunkte der Bildschirme und die Positionen der Infrarotstrahler vermessen werden. Zusätzlich zu den reellen Bildschirmen wurden im Weltmodell zusätzlich virtuell vergrößerte Bildschirme mit gleichem Zentrum und gleicher Orientierung angelegt, damit Blickmessungen, die einen Bildschirm leicht verpassen ebenfalls erfasst werden.

4.2.4.3 Kalibrierung der Blickrichtung

Zusätzlich zu der Erstellung des Weltmodells benötigt die Blickbewegungsmessung ein personalisiertes Modell für jede vermessene Person. Dieses Modell besteht aus Kalibrierungsparametern, die eine systematische Auslenkung zwischen der gemessenen Blickrichtung und der tatsächlichen Blickrichtung, also der Hauptrichtung der Informationsaufnahme, beschreiben. Für die Bestimmung dieser Auslenkung werden Messpunkte aufgenommen während die Versuchsperson vordefinierte Kalibrierpunkte fixiert. Die Kalibrierpunkte werden der Versuchsperson der Reihe nach angezeigt, sodass für jeden Bildschirm Messpunkte aufgenommen werden können. Die Abstände der Messpunkte zu den Kalibrierpunkten werden in einem Folgeschritt durch eine globale Optimierung minimiert.

Aufgrund der Verteilung der Programme für die Anzeige und die Blickbewegungsmessung muss erneut eine Interprozesskommunikation aufgebaut werden. Für diese Kommunikation bietet SmartEye eine JSON-RPC Schnittstelle an, mit der externe Programme das Weltmodell anfragen und den Kalibrierungsprozess fernsteuern können. Mit dieser Schnittstelle wurde ein Kalibrierprogramm geschrieben, das die Kalibrierpunkte der Reihe nach anzeigt. Dieses Programm wird durch Knöpfe an dem Stick des HOTAS-Systems ferngesteuert, damit die Versuchsperson das Fixieren des Kalibrierpunktes selbstständig bestätigen kann.

Obwohl die globale Optimierung der Kalibrierung die gesamten Abstände zwischen Messpunkten und Kalibrierpunkten minimiert, wurden für jeden Bildschirm individuelle systematische Abstände ermittelt. Damit diese Fehler ebenfalls beseitigt werden können, wurde ein zweiter Kalibrierschritt implementiert, bei dem erneut Messpunkte pro Bildschirm aufgenommen werden. Die damit ermittelten zweidimensionalen Abstände werden für eine bilineare Interpolation genutzt, die den kalibrierten Durchstoßpunkt erneut verzerrt.

4.2.4.4 Semantik der dargestellten Elemente

Im Rahmen der Forschungsarbeit an dem Kampfflugzeugsimulator wurde eine Softwarelösung entwickelt, die für jeden Bildschirm den Durchstoßpunkt des Blickes des menschlichen Operateurs mit der an dem Punkt dargestellten Information anreichern kann, um eine verbesserte Analyse des Blicks zu ermöglichen. Dies wird ermöglicht, indem die grafische Darstellung selbst geschrieben wurde und dabei jedem Element mitgeteilt wird welche Information es enthält, sodass daraufhin zu einem gegebenen Blickpunkt im Interface angefragt werden kann welche Information sich dort befindet. Abbildung 28 zeigt eine Darstellung des Interfaces. Auf der linken Seite der Abbildung sieht man das Interface so wie es der Pilot im Cockpit angezeigt bekommt und auf der rechten Seite wird dieselbe

Situation überlagert mit einer Kategorie der dargestellten Information. Hierbei sind blau überlagerte Interfaceelemente Knöpfe, die der Pilot betätigen kann, um mit dem Mensch-Maschine System zu interagieren. Mit einer orangenen Überlagerung werden Label gekennzeichnet, die statische Informationen anzeigen, wobei Label mit dynamisch veränderlichen Informationen, wie beispielsweise dem aktuell vorhandenen „Fuel on Board“, in grün überlagert werden. In Weiß dargestellte Interfaceelemente sind Objekte, der simulierten Welt wie Flugzeuge, Bodenabwehrstellungen oder selbstdefinierte logische Objekte.



Abbildung 28: Darstellung des zentralen Cockpitbildschirms links ohne und rechts mit Darstellung der semantischen Bedeutung

Die Anreicherung des Durchstoßpunktes mit der dargestellten Information erlaubt es Analysen 2ten Grades („*second-level analysis*“) des Blickverhaltens des Operators durchzuführen. Dies erlaubt ein differenziertes Bild über die Prozesse bei der Aufgabendurchführung, wie beispielsweise die Informationssammlung („*information-gathering*“), zu generieren (Friedrich et al. 2017). Für eine robustere Schätzung der semantischen Informationen wird die bestimmte Blickrichtung mit einer Abtastung der Anzeigeelemente im Umfeld des Durchstoßpunktes kombiniert. Somit besteht die semantische Bedeutung der Blickrichtung aus einer Verteilung der betrachteten Elemente statt eines einzelnen Datenpunktes, bei der jeder Abtastwert mit einer Normalverteilung auf Basis der Distanz zum Durchstoßpunkt gewichtet wird.

Für die Durchführung der Anreicherung ist es notwendig, dass jedes Anzeigeelement die eigene Bedeutung kennt. Für die korrekte Umsetzung dieses Wissens wurde eine Softwarearchitektur entwickelt und implementiert, die Programmierfehler bei der Erstellung von Anzeigeelementen vermeidet. Die Grundidee dieser Architektur besteht darin, dass jede Interaktion mit einem Anzeigeelement als ein ‚*Event*‘ angesehen wird und dieses Event dazu genutzt werden kann, die Darstellung des Elements an seine Funktion zu koppeln. Für diese Kopplung sieht die entwickelte Softwarearchitektur vor, dass jedes Anzeigeelement ein Objekt einer abstrakten Eventklasse erstellt und sowohl die Darstellung als auch die Funktion ausschließlich auf Basis dieses Objekts erzeugt werden kann. Somit muss das

erzeugte Objekt alle Informationen enthalten, die für beide Anwendungsfälle notwendig sind, weshalb jedes Anzeigeelement sein eigene Eventsubklasse benötigt.

Während der Ausführung des Programms wird auf Basis des Softwaremusters ‚*Visitor*‘ die Anzeige jedes Elements über den sogenannten ‚*StyleVisitor*‘ erzeugt. Das Visitormuster ermöglicht mit einem Zeiger auf ein Objekt der abstrakten Eventklasse die korrekte Subklasse des Events zu identifizieren und eine entsprechende Funktion aufzurufen. Jede Interaktion des Piloten, sowohl manuell als auch visuell, führt daraufhin zu der Auslösung des gleichen Subevents wodurch analog zum ‚*StyleVisitor*‘ ein ‚*EvidenceVisitor*‘ aufgerufen wird, der das Event in seine Bedeutung übersetzt. Im Falle einer manuellen Interaktion wird zusätzlich ein ‚*HandleVisitor*‘ ausgelöst, der im Scenario Programm die zugehörige Funktion verursacht.

4.2.5 Aktivitätsbestimmung

Auf Basis der semantischen Anreicherung der Interaktionen wird eine Echtzeitaktivitätsbestimmung während der Simulation durchgeführt, damit diese als Kontextwissen für adaptive Automationssysteme genutzt werden kann. Die Software beruht dabei auf der Aktivitätsbestimmung von (Honecker und Schulte 2017), wie sie bereits in Kapitel 3.2.2.2 beschrieben wurde. Für den Einsatz in der MUM-T Missionsapplikation mussten daran Änderungen vorgenommen werden, die im Folgenden beschrieben werden.

4.2.5.1 Adaption des Aufgabenmodells

Das Aufgabenmodell beschreibt die unterschiedlichen Aufgaben und deren Zusammenhänge in der jeweiligen Domäne. Einer der Kernpunkte der Übertragung der Aktivitätsbestimmung in eine neue Domäne war daher das Erstellen eines neuen Aufgabenmodells, von dem ein Auszug in Abbildung 29 dargestellt ist.

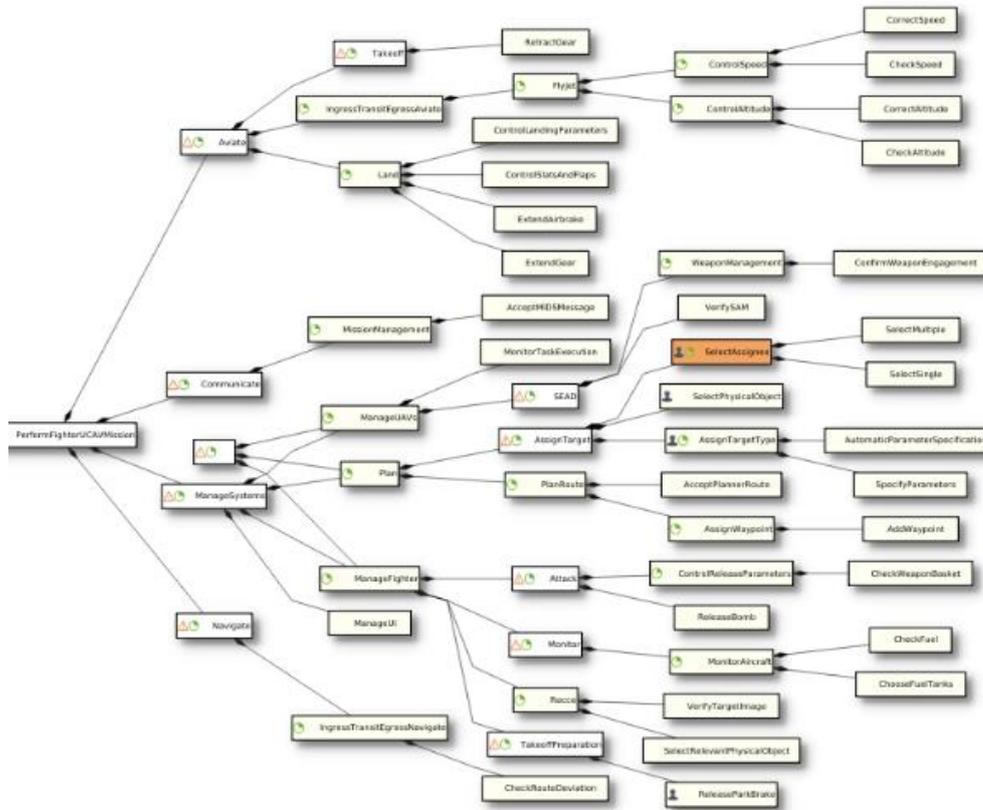


Abbildung 29: Auszug aus dem Aufgabenmodell

Die abgebildeten militärischen Kampfflugzeugmethoden benötigen eine hohe Interoperabilität unter den einzelnen befreundeten Einheiten, während die unbemannten Einheiten in der ursprünglichen Helikopterdomäne vorwiegend für unabhängige Aufklärungsaufgaben eingesetzt wurden. Diese zusätzliche Interoperabilität wurde in das Aufgabenmodell über eine dritte Beziehung zwischen den Aufgaben, der ‚logischen Abhängigkeit‘, modelliert. Zusätzlich bedeutet die Abhängigkeit der Aufgaben untereinander, dass neben den Aktivitäten des Menschen zusätzlich die Aktivitäten der unbemannten Flugzeuge teilweise modelliert werden müssen, damit die Abhängigkeitsbeziehungen korrekt abgebildet werden können. So hängt beispielsweise der automatisierte Wirkmitteleinsatz durch ein UAV von der visuellen Aufklärung des gleichen Objekts durch den Piloten ab. Diese muss bereits in der Aufgabenmodellierung berücksichtigt werden, indem auch der Wirkmitteleinsatz modelliert wird.

Identifikation der Aufgaben

Um das Aufgabenmodell zu erstellen, muss zunächst ermittelt werden, welche Aufgaben in der Zieldomäne enthalten sind. Diese Umgebungen sind domänenspezifisch und erfordern umfangreiches Wissen. Daher wurde über mehrere strukturierte Interviews mit Kampfflugzeugpiloten der Luftwaffe ein Grundwissensschatz über die notwendigen Tätigkeiten innerhalb der Kampfflugzeugdomäne aufgebaut. Diese Befragungen schaffen

einen Ausgangspunkt dafür, welche Aufgaben in der realen Welt ausgeführt werden. Bei simulierten Umgebungen müssen diese Aufgaben mit den Möglichkeiten des Simulators verglichen werden, um zu beurteilen, welche Aufgaben der realen Welt auf die simulierte Umgebung abgebildet werden. Zusätzlich beschäftigt sich die entwickelte Zielplattform mit der Untersuchung der Integration unbemannter Teammitglieder in zukünftige Missionen und damit verbundene neue Einsatzkonzepte, weshalb die Identifikation der relevanten Aufgaben erneut abstrahiert werden muss.

Bei der Einbindung der identifizierten Aufgaben in das Aufgabenmodell orientiert sich der Detaillierungsgrad der Aufgaben an den Anwendungsfällen eines verbundenen adaptiven Automationssystems, an der Aufgabenbewertung sowie an den Anforderungen der Arbeitsbeanspruchungsmodellierung. Das bedeutet, dass Aufgaben im Optimalfall so detailliert sein sollten wie die Warnungen, die generiert werden sollen, um z.B. Benachrichtigungen für bereits durchgeführte Aufgaben zu vermeiden. Hinzu kommt, dass komplexe Aufgaben unterschiedlich viel Arbeitslast verursachen, so dass die Modellierung der Arbeitslast nur für Aufgaben mit annähernd konstanter Arbeitslast möglich ist, was ebenfalls den Detaillierungsgrad der modellierten Aufgaben erhöht. Andererseits erschwert ein höherer Detaillierungsgrad die Unterscheidung der Aufgaben voneinander, wenn sie ähnliche Interaktionsmuster aufweisen. So sind beispielsweise Änderungen von Systemparametern auf niedriger Ebene, wie das Deaktivieren oder Aktivieren einer Parkbremse, nicht voneinander zu unterscheiden, wenn der aktuelle Zustand der Parkbremse nicht bekannt ist. Daher ist es notwendig, entweder diese Information als zusätzliche Kontextparameter zu generieren oder den Detaillierungsgrad der Aufgabenbewertung so zu reduzieren, dass lediglich die Änderung der Parkbremse erkannt wird, was ausreicht, um Systemmeldungen über einen falschen Parkbremszustand zu verhindern. Insgesamt zeigt dies, dass der Detaillierungsgrad der Aufgaben so hoch wie nötig, aber so niedrig wie möglich sein sollte.

Zuordnung von Evidenzen

Mit einem Aufgabenmodell der Zielumgebung ist der nächste Schritt die Ermöglichung der Aufgabenbewertung in Echtzeit durch die Zuordnung von Evidenzen (beschrieben in Kapitel 3.2.2.2) zu den entsprechenden Aufgaben. Dieser Satz von Evidenzen besteht in erster Linie aus den Interaktionen, die zur Durchführung der jeweiligen Aufgabe notwendig sind. Dabei sollen visuelle Interaktionen, die durch den Einsatz eines Blickverfolgers erfasst werden, eine geringere Überzeugung erzeugen als manuelle Interaktionen. Dies berücksichtigt unbewusste Blickinteraktionen mit dem System, wie z.B. das Anstarren eines Punktes, um nachzudenken, ohne visuelle Informationen wahrzunehmen, während manuelle Interaktionen in der Regel absichtsgesteuert sind.

Darüber hinaus sind Informationen über den aktuellen Zustand des automatisierten Systems als zusätzliche Evidenz notwendig, um verschiedene Aufgaben mit ähnlichen Interaktionsmustern zu unterscheiden. So kann beispielsweise die Bewegung in der Steuerung eines Flugzeugs als Form der Eingabe nicht näher unterschieden werden, aber ein unabhängiges Verständnis der aktuellen Flugphase ermöglicht es, einen Unterschied zwischen einer Flugaufgabe und einer Überprüfung des Flugzeugs am Boden zu erkennen. Darüber hinaus bietet diese Art des Verständnisses die Möglichkeit, die Analyse für eine große Teilmenge von Aufgaben zu deaktivieren (durch Zuweisung eines Zweifelswertes von 1 für eine bestimmte Evidenz), was einerseits das Verständnis der Benutzer für die Aufgabenbewertung erleichtert und andererseits den Bewertungsprozess für große Modelle drastisch beschleunigen kann, indem bereits unmögliche Aufgaben nicht weiter untersucht werden.

Definition der Beanspruchungsparameter

Der letzte Schritt bei der Erstellung des Aufgabenmodells für eine kontextreiche Ermittlung des mentalen Zustands ist die Definition der Arbeitsbeanspruchungsanforderungen gemäß der bereits beschriebenen ‚Multiple Resource Theory‘ (siehe Kapitel 3.1.4). Ähnlich wie bei den Evidenzen muss jede Aufgabe mit einer entsprechenden Anforderung verbunden werden. Die Beanspruchungsanforderungen wurden ähnlich wie die Aufgaben selbst auf Basis von Experteninterviews bestimmt. Eine Maßnahme dabei war es bestimmte Mehraufgabensituationen von den Experten direkt bewerten und dabei die einzelnen Parameter genauer beschreiben zu lassen. Diese Vorgehensweise stellt die relevanten Informationen in einer Situation genauer dar und lässt dadurch Rückschlüsse auf die Beanspruchung der einzelnen Aufgaben zu. Abbildung 30 zeigt ein Beispiel für die Anforderungen bei die Aufgabe "Planen".

Properties

Name

Categories **Inherited Cat's**

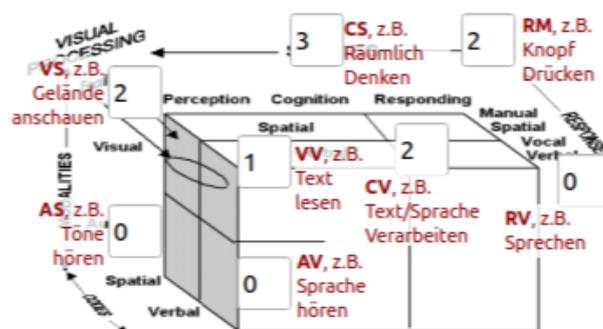


Abbildung 30: Beanspruchungsdefinition der Planaufgabe, wie in (Mund und Schulte 2018)

Die Aufgabe ist charakterisiert durch die Bewertung von relativen Positionen der eigenen Kräfte sowohl zueinander, als auch zu dem gewünschten Zielobjekt. Die Analyse dieser Positionen bedarf der Wahrnehmung der Informationen mit Hilfe der taktischen Kartendarstellung, sowie deren Verarbeitung. Andererseits stellt diese Art von Informationen die Grundlage der Tätigkeit eines Kampfflugzeugpiloten in einer Führungsposition dar, wodurch ihm diese Abstände in der Regel bewusst sind. Diese gegensätzlichen Rahmenbedingungen führen zu insgesamt niedrigen Beanspruchungsniveaus von zwei für den visuell räumlichen Kanal und drei für den kognitiv räumlichen Verarbeitungskanal auf einer Skala von eins bis zehn. Bei dem visuell verbalen und dem kognitiv verbalen Kanal existiert eine ähnliche Beanspruchungssituation. Die zugehörigen Informationen sind die verfügbaren Ressourcen der Teammitglieder, wie beispielsweise deren Sensoren und Wirkmittel. Auch hierbei handelt es sich um vorwiegend bekannte Informationen, die sich zusätzlich wenig verändern und in einer Anzeige permanent dargestellt werden, wodurch die zugehörige Beanspruchung etwas niedriger ist, als die der räumlichen Kanäle. Die auditiven Kanäle werden in derzeitigen Kampfflugzeugmissionen stark beansprucht, da die Auftragszuweisung, sowie die Abstimmung unter den Teammitgliedern, verstärkt über die Funkschnittstelle stattfindet. In der Simulationsumgebung wurde diese Kommunikation allerdings aufgrund der Änderung hin zu unbemannten Kampfflugzeugen nicht abgebildet, weshalb die Beanspruchung der Kanäle jeweils mit null von zehn Punkten bewertet wurde.

4.2.5.2 Anpassung der Prozesskette

Neben dem Aufgabenmodell mussten zusätzlich Änderungen an der Softwareprozesskette vorgenommen werden, damit sie den Anforderungen der neuen Domäne entspricht. Die einzelnen Module der Prozesskette sind dabei erhalten geblieben und entsprechen weiterhin dem Blockschaltbild aus Abbildung 18. Allerdings musste der Inhalt der Module an die veränderte Umgebung angepasst werden. Die in Kapitel 4.2.4.1 beschriebene semantische Analyse bildet das ‚Evidence Generation‘ und wurde von Grund auf neu für diesen Anwendungsfall entwickelt. Dies war notwendig, weil die Semantik an die Anzeige der Mensch-Maschine-Schnittstelle angepasst werden musste und die Schnittstelle des Kampfflugzeugs von Grund auf neu entwickelt wurde.

Das Modul ‚Task Assessment‘ wurde ebenfalls angepasst. Wie zuvor beschrieben stehen in der Kampfflugzeugdomäne die Wechselwirkungen zwischen Aufgaben stärker im Fokus. Gleichzeitig befinden sich in den modellierten Zielgebieten häufig viele Objekte des gleichen Typs. Daher ist es nicht ausreichend eine Aufgabe unabhängig von dem betroffenen Objekt zu modellieren, da sonst die Abhängigkeiten der Aufgaben nicht repräsentiert werden kann. Um dieser geänderten Anforderung gerecht zu werden, kann in der angepassten Variante der Aktivitätsbestimmung die gleiche Aufgabe mehrmals existieren, und zwar

einmal pro Objekt, auf der die Aufgabe ausgeführt werden kann. Mit dieser Änderung der Aufgabenbeschreibung konnte der Dempster-Shafer Inferenzalgorithmus unverändert genutzt werden, um die Bestimmung der Aktivitäten durchzuführen. Zusätzlich konnte auch das Modul der ‚Workload Estimation‘, sowie das Ergebnis der Aktivitätsbestimmung direkt übernommen werden und bedurfte keiner Anpassungen.

4.3 Kontextfreie Ermittlungsmethoden

Für die Untersuchung der kontextfreien Ermittlungsmethoden wurden zwei Untersuchungsmethoden ausgewählt: die Messung von kardiovaskulären Parametern mit Hilfe eines Elektrokardiografen und die Messung des Pupillendurchmessers durch ein Kamerabasiertes Blickbewegungsmesssystem. Die Herzdaten stellen einen weit verbreiteten Indikator für die Beanspruchung dar und versprechen eine robuste Beanspruchungsschätzung im Vergleich zu Messungen auf Basis der unterschiedlichen Hirnareale. Der Pupillendurchmesser hingegen stellt ein Maß dar, welches intrusionsfrei eingerüstet werden kann und deshalb zusätzliche Vorteile mit sich bringt.

4.3.1 Pupillenbasiertes Experiment

Die Zielsetzung dieses Experiments bestand darin unterschiedliche Arten von Beanspruchung zu stimulieren, um zu evaluieren inwiefern diese Arten von Beanspruchung unterschieden werden können. Die Messung basiert hierbei auf „Haar-Like Features“ (Lienhart und Maydt 2002), die auf Basis der Pupillenweite berechnet werden. Basierend auf gleichgearteten Features konnten (Xu et al. 2011) einen Boosting (siehe (Freund et al. 1999)) Klassifikator trainieren, der unterschiedliche kognitive Beanspruchungsniveaus voneinander unterscheiden kann und dabei robust gegenüber Veränderungen in der Umgebungsbeleuchtung ist. Des Weiteren können laut (Mathôt 2018) unterschiedliche Formen von psychosensorischen Pupillenreaktionen erkannt werden. Mathôt beschreibt hierbei vorwiegend die Unterscheidung zwischen der orientierenden Reaktion, bei der unerwartete Informationsdarstellungen zu einer visuellen Reaktion führen, und der mentalen Anstrengung, die bei der Ausübung einer kognitiven Aufgabe auftritt. Die Fragestellung des Experiments bestand darin herauszufinden inwieweit diese unterschiedlichen Reaktionen der Pupille dazu genutzt werden können Rückschlüsse auf die zugrundeliegende Form von Beanspruchung geschlossen werden können.

4.3.1.1 Experimentaldesign

Um die Zielsetzung des Experiments zu erreichen, wurden vier voneinander unabhängige Teilerperimente gestaltet. Jedes der Teilerperimente wurde dabei so konzipiert, dass es sich möglichst gut auf eine der vier Beanspruchungsdimensionen des in Kapitel 3.1.2 vorgestellten VACP Modells abbilden lässt, weshalb dieses Experiment nicht die Zieldomäne abbildet. Jedes dieser Teilerperimente wurde in zwei unterschiedlichen Schwierigkeitsstufen durchgeführt. Jedes der Experimente wurde fünf Minuten lang durchgeführt.

Visuell

Das visuelle Teilerperiment ist eine Abwandlung des Stroop Tests (Stroop 1935). Den Versuchsteilnehmern wurde eine Textgleichung auf dem zentralen Cockpitbildschirm angezeigt. Die linke Seite der Gleichung ist ein Farbwort, dessen Bedeutung und Anzeigefarbe sich unterscheiden. Abbildung 31 stellt eine beispielhafte Textgleichung dar in der das Farbwort mit der Bedeutung „Cyan“ in der Anzeigefarbe „rot“ visualisiert wird. Die rechte Seite der Textgleichung beinhaltet eine Fragestellung zu dem dargestellten Farbwort. Bei dieser Fragestellung wird entweder die Bedeutung (z.B. „Text Cyan“ wie in Abbildung 31 oben) oder die Anzeigefarbe (z.B. „Color Cyan“ wie in Abbildung 31 unten) des Farbworts abgefragt. Die Aufgabe der Versuchsteilnehmer war es, jede wahre Textgleichung mit einem Knopfdruck zu bestätigen, wobei bei einer falschen Textgleichung nichts zu tun war. Je nach Schwierigkeit wurde eine neue zufällig generierte Aufgabe alle 2,5 bzw. 1,5 Sekunden angezeigt.

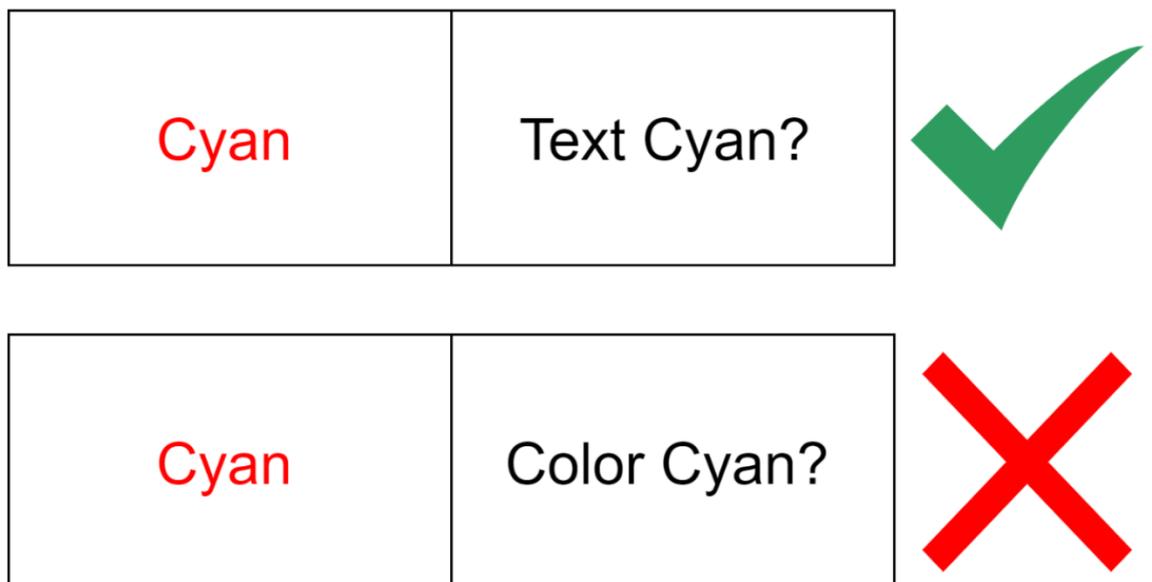


Abbildung 31: Beispielhafte Textgleichungen im Pupillenbasierten Experiment

Auditiv

Bei dem auditivem Telexperiment wurden den Versuchsteilnehmern Buchstaben nach dem „NATO phonetic alphabet“ von einer Sprachsynthesoftware vorgelesen. Für die Tonausgabe erhielten die Versuchsteilnehmer einen Kopfhörer. Die Aufgabe der Versuchsteilnehmer war ein 1-Back Test. Wann immer der aktuelle Buchstabe dem vorherigen Buchstaben entspricht, mussten die Versuchsteilnehmer dies mit einem Knopfdruck bestätigen, wie in Abbildung 32 dargestellt. Auch hier wurde ein neuer Buchstabe je nach Schwierigkeit entweder alle 2,5 oder 1,5 Sekunden vorgelesen.

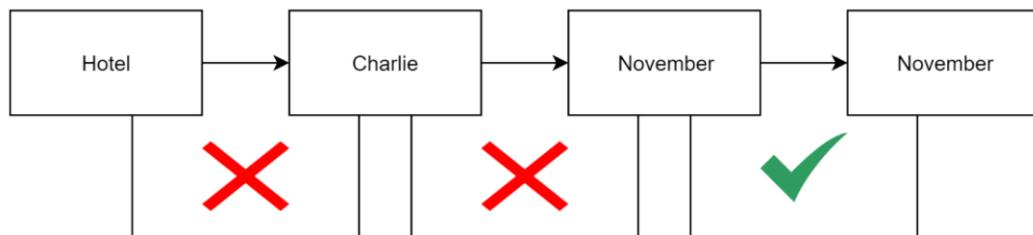


Abbildung 32: Visuelle Darstellung des Auditiven Telexperiments

Kognitiv

Das Kognitive Telexperiment bestand aus Additions- oder Subtraktionsaufgaben von zwei zweistelligen Zahlen. Die Aufgaben wurden zufällig generiert und auf dem zentralen Cockpitbildschirm angezeigt. Zusätzlich wurde die korrekte Antwort zusammen mit drei weiteren Antwortmöglichkeiten dem Versuchsteilnehmer angezeigt, wobei bei jeder Aufgabe die linke Antwort farblich hervorgehoben wurde. Abbildung 33 visualisiert die beschriebene Darstellung. Die weiteren Antwortmöglichkeiten wurden hierbei zufällig generiert, indem eine 1-Stellige oder kleine 2-Stellige Zahl auf die korrekte Antwort addiert oder subtrahiert wurde. Die Versuchsteilnehmer mussten die korrekte Antwort aus den vier Antwortmöglichkeiten auswählen, indem sie mit dem HOTAS-System im Cockpit per Knopfdruck die Markierung nach rechts bzw. links verschieben und mit einem weiteren Knopfdruck die korrekte Auswahl bestätigen. In der einfachen Stufe wurde eine neue Aufgabe alle 7,5 Sekunden und in der schwierigen Stufe alle 5 Sekunden eingespielt.

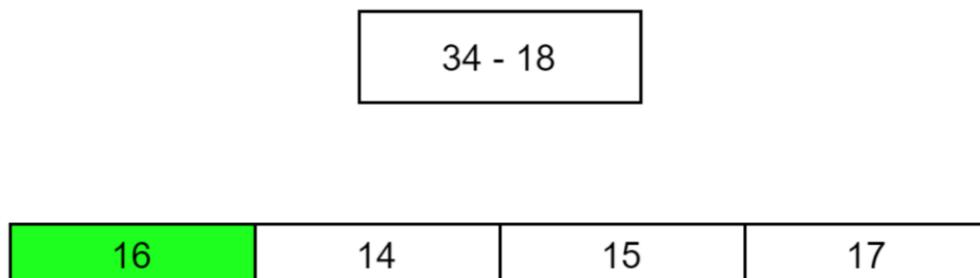


Abbildung 33: Darstellung einer Subtraktionsaufgabe im kognitiven Telexperiment

Psychomotorisch

Die Aufgabe bei dem psychomotorischen Telexperiment war es eine Kugel auf einem Balken zu balancieren. Hierbei konnten die Versuchsteilnehmer mit den Pedalen im Simulator Cockpit den Balken um das Zentrum herum rotieren und somit die Geschwindigkeitsänderung der Kugel zu manipulieren (siehe Abbildung 34). Sobald die Kugel den Rand des Balkens erreicht hat, wurde sie gestoppt und konnte nur eine Geschwindigkeit in die entgegengesetzte Richtung aufbauen. Die Schwierigkeit in diesem Telexperiment wurde manipuliert, indem die simulierte Trägheit der Kugel verändert wurde. Auf der höheren Schwierigkeitsstufe wurde diese Trägheit verkleinert und somit die winkelabhängige Geschwindigkeitsänderung der Kugel erhöht.

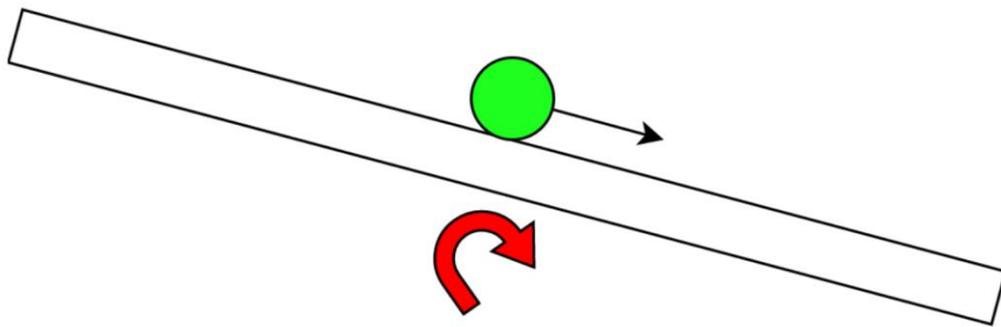


Abbildung 34: Darstellung der Balancieraufgabe des psychomotorischen Telexperiments

4.3.1.2 Einordnung der Telexperimente

Um sicher zu stellen, dass die jeweiligen Telexperimente der entsprechenden VACP-Dimension entsprechen, wurden die Aufgaben innerhalb des Experiments genauer analysiert. Hierzu haben (Bierbaum et al.) Deskriptoren aufgestellt, um die Schwierigkeiten der einzelnen VACP-Dimensionen greifbar zu machen, beispielsweise ist hierbei das Registrieren eines Geräuschs die einfachste auditive Aufgabe. Mit Hilfe dieser Deskriptoren wurde für die einzelnen Telexperimente ermittelt, wie die Ausgeführten Aufgaben die einzelnen VACP-Dimensionen Beanspruchen. Als Beanspruchungswert wurden dazu die Umfrageergebnisse aus (SCHUCK 1996) benutzt. Die Beanspruchungswerte wurden dabei durch Piloten des CP140 Aurora den einzelnen Deskriptoren zugeordnet. Abbildung 35 stellt diese Bewertungen grafisch dar.

Experiment/ Komponente	Visuelles	Kognitives	Auditives	Psychomotorisches
Visual	Discriminate (Detect Visual Differences) = 6.5	Read (Symbol) = 3.1	/ = 0	Register/ Detect (Detect Occurrence of Image) = 1.0
Cognitive	Automatic (Simple Association) = 1.0	Estimation, Calculation, Conversion = 4.0	Automatic (Simple Association) = 1.0	Automatic (Simple Association) = 1.0
Auditory	/ = 0	/ = 0	Interpret Semantic Content (Speech) = 4.8	/ = 0
Psychomotor	Discrete Actuation (Button, Toggle, Trigger) = 1.3	Discrete Actuation (Button, Toggle, Trigger) = 1.3	Discrete Actuation (Button, Toggle, Trigger) = 1.3	Continuous Adjustive (Flight Control, Sensor Control) = 3.8

Abbildung 35: Beanspruchungen der VACP-Dimension für die Telexperimente im Pupillenbasierten Experiment (Spalte = Art des Experiments, Zeile = VACP-Dimension)

Anhand dieser Bewertung ist für jedes Experiment die Beanspruchung in der geplanten VACP-Dimension am höchsten. Allerdings ist die kognitive Beanspruchung innerhalb des kognitiven Telexperiments nur leicht höher als die visuelle Beanspruchung.

4.3.1.3 Durchführung

Das Experiment wurde mit acht Piloten der Luftwaffe durchgeführt. Die Piloten waren alle männlich und haben eine komplette Kampfpilotenausbildung durchlaufen. Jeweils eine Hälfte der Piloten wurde hierbei für den Tornado beziehungsweise den Eurofighter ausgebildet. Die Demographie und die Erfahrung der Piloten wird in Tabelle 4 dargestellt.

Tabelle 4: Demographische Daten und Erfahrung der Versuchsteilnehmer

Alter in Jahren	Ausgebildete Flugsysteme	Flugerfahrung in Stunden	Einsatzerfahrung in Stunden
35	Tornado	1000	210
32	Tornado	1200	200
39	Tornado	1500	120
35	Tornado	1000	100
51	F-4F MiG-29 Eurofighter	3500	-
38	F-4F Eurofighter	1160	-
30	Eurofighter	1000	50
30	Eurofighter	480	-

Zu Beginn des Experiments wurden die Telexperimente vorgestellt und das Blickbewegungsmesssystem auf den jeweiligen Versuchsteilnehmer kalibriert. Während der Durchführung wurde den Versuchsteilnehmern vor jedem Telexperiment erneut das Telexperiment vorgestellt und die notwendigen Bedienelemente dargelegt. Während der

Experimente wurde die Pupillenweite der Probanden mithilfe des Blickbewegungsmesssystems aufgenommen.

Auf den resultierenden Pupillendaten wurden daraufhin Haar-Like Features ermittelt. Die Erkennung basiert hierbei auf der „Task Evoked Pupillary Response“ (Beatty 1982), also der Reaktion der Pupille auf die Informationsverarbeitung einer Aufgabe. Deshalb werden die Haar-Like Features in der Zeit (1,5 Sekunden) kurz nach Auftreten einer neuen Aufgabe berechnet. Das psychomotorische Experiment stellt hierbei eine kontinuierliche Aufgabendurchführung dar und die Berechnung der Haar Like Features wurde in 1,5 Sekunden lange Fenster diskretisiert.

Wie in Abbildung 36 dargestellt, werden für diese Zeitfenster systematisch Schablonen generiert, die aus einer Dauer und einem Zeitpunkt relativ zum Beginn des Zeitfensters bestehen. Die Schablone wird weiterhin in zwei Hälften geteilt und für jede der beiden Hälften wird die Summe der Pupillenweiten berechnet. Für jede der Schablonen ergibt die Differenz der beiden Hälften einen Haar-Like Feature Wert. Die hierdurch erstellten Features zusammen mit den Absolutwerten der Pupillenweiten bilden die Grundlage für das Training von mehreren Verfahren des überwachten maschinellen Lernens (siehe (Kotsiantis et al. 2007)).

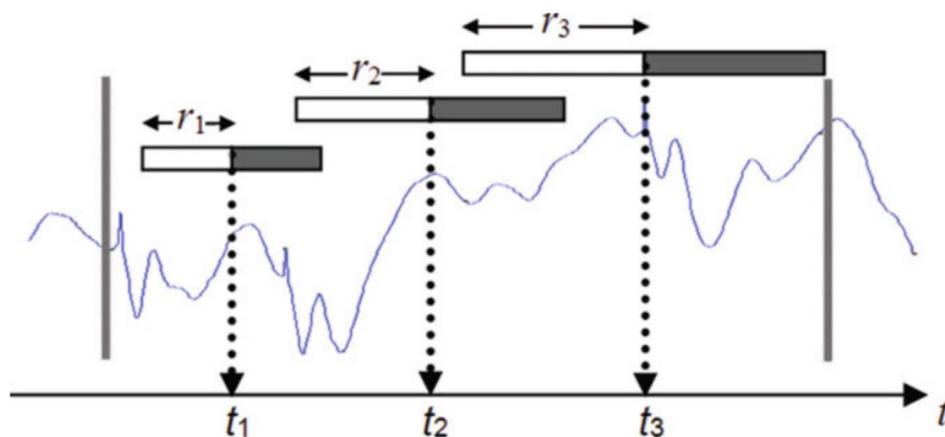


Abbildung 36: Berechnung von Haar-Like Features auf der Pupillenweite nach (Chen et al. 2016)

Für das Training wurden mithilfe der C++ Bibliothek OpenCV (Bradski und Kaehler 2000) mehrere unterschiedliche Klassifikatoren trainiert und mit einander verglichen. Die hierbei verwendeten Algorithmen sind die OpenCV Implementationen der Folgenden Klassifikatoren:

- AdaBoost mit Decision Stumps (Schapire 2013)
- K-Nearest Neighbor (Cover und Hart 1967)
- Random Forest (Biau und Scornet 2016)
- Support Vector Machine (Cortes und Vapnik 1995)

- Neuronale Netze (Ridella et al. 1997)

Die entstehenden Klassifikatoren werden zur Testphase mit den Haar-Like Features der Pupillenweite befüllt, um darauf aufbauend ein Klassifikationsergebnis zu generieren. Die Berechnung der Features basiert dabei auf einem ‚*Sliding-Window approach*‘ über der Zeit, anstatt erneut einem Auslösen auf Basis des Wissens über die Anzeige einer neuen Aufgabe. Dadurch wird simuliert, dass zu einem Testzeitpunkt das Wissen über das Auftreten neuer Aufgaben nicht vorhanden ist.

4.3.1.4 Ergebnisse

Die trainierten Klassifikatoren wurden mithilfe einer Leave-one-out-cross-validation (LOOCV) (Wong 2015) auf ihre Klassifikationsgüte evaluiert. Dazu wurden auf Basis der Daten von sieben Piloten die unterschiedlichen Klassifikationsalgorithmen trainiert und die resultierenden Klassifikatoren auf den Daten des verbleibenden Piloten getestet. Dieses gängige Verfahren im Bereich des maschinellen Lernens soll das als *Overfitting* bekannte Problem verhindern, bei dem ein gutes Klassifikationsergebnis nur dadurch erzielt wird, dass die Testdaten bereits von dem Klassifikationsalgorithmus gesehen wurden. Bei der beschriebenen Cross-Validation konnten Neuronale Netze durchweg ein besseres Klassifikationsergebnis erzeugen als die anderen Klassifikationsalgorithmen. Tabelle 5 stellt auszugsweise den Vergleich der Klassifikatoren für das visuelle Telexperiment als durchschnittliche Cross-Validation über alle Piloten dar. Hierbei ist die Klassifikationsgenauigkeit des Neuronalen Netzes (MLP) bei dem jeweiligen Testdatenset mehr als doppelt so hoch als bei den anderen Klassifikationsalgorithmen. Daher wird ab diesem Punkt nur noch der Klassifikator auf Basis des neuronalen Netzes berücksichtigt.

Tabelle 5: Vergleich der Klassifikationsalgorithmen im visuellen Experiment

Experiment	Algorithmus	Genauigkeit	Durchschnitt
Visual	Ada	Train	79.3297
		Test	35.910075
	Knear	Train	82.2957625
		Test	31.01485
	Rtree	Train	79.786375
Test		35.1728125	
SVM	Train	71.9508	
	Test	36.008975	
MLP	Train	99.4993875	
	Test	74.396675	

Für alle vier Telexperimente wurde ein eigener one-vs-all Klassifikator trainiert, der das Auftreten des jeweiligen Experiments erkennen soll. Mit der bereits beschriebenen Methode wurde für jedes Experiment die Klassifikationsgenauigkeit bestimmt. Das Ergebnis wird in Tabelle 6 für die jeweils besten Klassifikatoren dargestellt (alle Klassifikatoren basieren auf dem Neuronalen Netz als Klassifikationsmechanismus). Es zeigt sich, dass die drei Telexperimente mit klar definierten Aufgabenreizen gut erkannt werden können, während das psychomotorische Experiment lediglich in 64.9% der Fälle erkannt werden konnte.

Tabelle 6: Vergleich von Trainings- und Testgenauigkeit der vier Klassifikatoren

	Visual-Classifier	Auditory-Classifier	Cognitive-Classifier	Psychomotor-Classifier
Training-Accuracy	99.5	99.5	99.5	99.5
LOOCV-Accuracy	75.0	82.4	89.2	64.9

Des Weiteren wurden die Telexperimente mit einer 2ten Personengruppe, bestehend aus Studenten und Mitarbeitern des Instituts für Flugsysteme erneut durchgeführt, um ein zusätzliches Validierungsdatenset zu generieren, mit dem weitergehend gezeigt werden kann, dass sich das Klassifikationsergebnis generalisieren lässt. Tabelle 7 stellt das Ergebnis der Evaluation der Cross-Validation, dem Ergebnis der Testung mit der zweiten Personengruppe gegenüber. Die Klassifikationsalgorithmen wurden dabei auf den Daten aller acht Piloten trainiert und das Ergebnis daraufhin auf den Daten der zweiten Personengruppe evaluiert.

Tabelle 7: Genauigkeit der vier Klassifikatoren auf dem Validierungsdatenset (2. Versuchsteilnehmergruppe)

	Visual-Classifier	Auditory-Classifier	Cognitive-Classifier	Psychomotor-Classifier
LOOCV-Accuracy	75.0	82.4	89.2	64.9
2nd Group Accuracy	75.8	81.6	88.8	74.2

Die Güte der Klassifikation auf Daten der zweiten Personengruppe ist hierbei dem Klassifikationsergebnis der Cross-Validation sehr ähnlich. Einzig der Klassifikator für das psychomotorische Telexperiment weicht von dem ursprünglichen Ergebnis ab und liefert dabei sogar ein um nahezu zehn Prozent genaueres Resultat.

Zusätzlich zu der Unterscheidung der unterschiedlichen Telexperimente wurden weitere Klassifikatoren trainiert, die Aussagen darüber generieren sollen, welche Schwierigkeit des

Experiments vorliegt. Die Genauigkeit der Klassifikation lag dabei allerdings für alle vier Experimente bei ca. 50%. Somit war die Schwierigkeitsklassifikation nicht erfolgreich und wurde verworfen.

4.3.1.5 Bewertung der Messmethode

Im Rahmen des Experiments konnte gezeigt werden, dass der Pupillendurchmesser einen Indikator für Beanspruchung darstellt. Hierbei konnte nicht nur Beanspruchung als solche erkannt, sondern auch unterschiedliche Arten von Beanspruchung voneinander unterschieden werden. Dabei ist zu beachten, dass die ermittelte Beanspruchung durch die Informationsaufnahme der Aufgabenstellung erkannt wird und somit die Beanspruchung im Rahmen von der Informationsverarbeitung nicht erfasst werden kann. Des Weiteren muss in weiteren Experimenten noch geklärt werden, ob die unterschiedlichen Ausschläge der einzelnen Klassifikatoren tatsächlich unterschiedliche Beanspruchungsdimensionen repräsentieren oder lediglich eine Unterscheidung der Experimente zueinander. In beiden Fällen stellt die Klassifikation eine wertvolle Information für die adaptive Automation dar, aber die Allgemeingültigkeit der generierten Aussage ist weitaus höher, wenn Beanspruchungsdimensionen unterschieden werden können.

Robustheit

Die Messung des Pupillendurchmessers kann durch kamerabasierte Blickbewegungsmesssysteme zuverlässig automatisiert ermittelt werden. Außerdem kann der häufig beschriebenen Lichtempfindlichkeit der Messmethode durch die Verwendung von Haar-Like Features entgegengewirkt werden, da die Messung ausschließlich Veränderungen des Pupillendurchmessers berücksichtigt und somit nur durch hochfrequente Lichtverhältnisänderungen beeinflusst werden kann. Damit ist die Robustheit der Messmethode insgesamt **sehr hoch**.

Sensitivität

Im Rahmen des durchgeführten Experiments konnte im Gegensatz zu (Chen et al. 2016) keine Unterscheidung von Beanspruchungsniveaus erreicht werden. Außerdem stellt die beschriebene Klassifikation von drei Beanspruchungsstufen keine eindeutige Identifikation der Notwendigkeit von zusätzlichen Unterstützungsfunktionen dar. Somit ist die Sensitivität der Messmethode **niedrig**.

Diagnostizität

Das Identifizieren der einzelnen Experimente allein aufgrund der Pupillendaten, obwohl gerade die ersten drei Experimente sehr ähnlich strukturiert waren, spricht dafür, dass die vier Dimensionen des VACP-Modells voneinander unterschieden werden können. Auch wenn diese Unterscheidung noch in weiteren Experimenten verifiziert werden muss, können

verschiedene Beanspruchungsprofile voneinander abgegrenzt werden. Dadurch weist die Messmethode eine **sehr hohe** Diagnostizität auf.

Individualität

Aufgrund seiner Charakteristik als direkte psychophysiologische Messmethode ist das pupillenbasierte Verfahren auf Basis des Pupillendurchmessers sehr individuell. Ausschließlich beanspruchende Situationen werden als solche klassifiziert. Zusätzlich benötigt das Verfahren keine Kalibrierung durch eine Baseline Messung, erneut da ausschließlich Änderungen der Pupillenweite genutzt werden. Auch für die Erkennung der Augen und der Ermittlung der Pupillenweite ist keine Kalibrierung notwendig. Insgesamt ist die Individualität der Messung **sehr hoch**.

Generalisierbarkeit

Ähnlich wie die elektrokardiografbasierte Messung benötigt die Beanspruchungsermittlung auf Basis des Pupillendurchmessers keinerlei zusätzliche Kontextinformationen, um eine Aussage zu generieren. Zusätzliche Informationen über das Auftreten von Ereignissen kann hingegen als zusätzliche Kontextinformation genutzt werden, um die Task Evoked Pupillary Response genauer zu untersuchen und somit das ermittelte Ergebnis besser einzuordnen. Nichtsdestotrotz kann die Bestimmung unabhängig der Rahmenbedingungen eingesetzt werden und etwaige Veränderungen der Umgebungsbeleuchtung werden durch die Ermittlungsmethode rausgefiltert. Dadurch ist die Generalisierbarkeit der Ermittlungsmethode **sehr hoch**.

Direkte Anwendbarkeit

Die Beanspruchungsermittlung auf Basis des Pupillendurchmessers eignet sich zur Unterscheidung mehrerer Arten von Beanspruchung. Auch wenn eine Zuordnung der Arten von Beanspruchung zu den VACP Dimensionen weiterhin getestet werden muss können entsprechend gelernte Klassifikatoren einen Indikator generieren welche Art von Aufgabe für die ausgelöste Beanspruchung verantwortlich ist. Somit können erste Einschätzungen generiert werden welche Aufgaben Unterstützung benötigen könnten. Demgegenüber generiert die Beanspruchungsbestimmung keine genaue Einschätzung für die Behebung der Beanspruchungsursache und somit ist die Direkte Anwendbarkeit trotzdem **niedrig**.

4.3.2 Elektrokardiografbasiertes Experiment

Ziel des Experiments war es zu untersuchen, inwiefern ein Elektrokardiograf eingesetzt werden kann um in nahezu Echtzeit Aussagen über den Beanspruchungszustand zu generieren. Wie bereits in Kapitel 3.2.1.1 erwähnt besteht ein erwiesener Zusammenhang zwischen mentaler Beanspruchung und der Herzrate, sowie der Herzratenvariabilität, eines

Operateurs. Allerdings basieren solche Analysen in der Regel auf einem Zeitfenster von mehreren Minuten. Zusätzlich konnten (Li et al. 2019) zeigen, dass im Bereich der Krankheitsdiagnose über Analysen der Herzratenvariabilität nur zuverlässige Aussagen generiert werden können, wenn sich das zugrundeliegende Zeitintervall über mehrere Minuten erstreckt.

Neben dem Elektrokardiogrammen wurde für dieses Experiment auch das Messen des Hautleitwiderstands in Erwägung gezogen. Aufgrund der notwendigen manuellen Interaktionen mit dem Kampfflugzeugcockpit mussten diese Messungen allerdings verworfen werden, da Sensoren für die Hautleitwiderstandsmessung an den Fingern der Probanden angebracht werden müssen.

4.3.2.1 Experimentaldesign

Die Zielsetzung bestand daraus mehrere unterschiedliche Schwierigkeitsstufen bereits im Experimentaldesign abzubilden, um die Selektivität der Messmethode genauer bewerten zu können. Aus diesem Grund wurde das Experiment in fünf voneinander möglichst separierbare Phasen unterteilt. Die Hauptaufgabe im Experiment bestand daraus eine vorgefertigte Route mit dem simulierten Kampffjet abzufliegen. Hierbei gab es eine Höhen- sowie eine Geschwindigkeitsvorgabe, die bestmöglich eingehalten werden sollten. Als zusätzliche Nebenaufgaben mussten die Probanden Kopfrechenaufgaben lösen, die ihnen diktiert wurden und über das taktische Display Aufgaben an drei unbemannte Kampfflugzeuge zuweisen (*Delegieren*). Diese drei Aufgabentypen wurden in fünf unterschiedlichen Schwierigkeitsstufen kombiniert wie in Tabelle 8 dargestellt wird. Zielsetzung dieser Schwierigkeitsstufen war hierbei ein klarer Anstieg von jeder Schwierigkeitsstufe im Vergleich zu der vorherigen.

Tabelle 8: Geplante Schwierigkeitsstufen des Elektrokardiografbasierten Experiments

Phase	Tasks
Schwierigkeit I	Fliegen
Schwierigkeit II	Fliegen, Delegation
Schwierigkeit III	Fliegen, Kopfrechnen
Schwierigkeit IV	Fliegen, Delegation, Kopfrechnen
Schwierigkeit V	Tiefflug, Delegation, Kopfrechnen

Kopfrechnen

Die zu bearbeitenden Kopfrechenaufgaben wurden hierbei während der entsprechenden Flugphasen vorgelesen und bestanden aus zweistelligen Additions- und Multiplikationsaufgaben mit jeweils zwei Rechenschritten wie:

$$\frac{49}{7} + 4$$

Das Ergebnis wurde von den Probanden verbal zurückgegeben. Bei einer falschen Antwort wurde dieselbe Aufgabe erneut vorgelesen, bis der Proband korrekt geantwortet hat.

Delegation

Bei der Delegationsaufgabe mussten die Probanden mit dem zentralen Touchbildschirm im Cockpit taktische Befehle an drei kooperierende unbemannte Flugzeuge vergeben. Hierzu gab es vier Arten von Objekten in der taktischen Kartendarstellung und jeweils eine zugehörige Aufgabe, die zugewiesen werden sollte. Bei Wegpunkten sollte eine „Fly-Over“ Aufgabe zugewiesen, Gebäude sollten über eine „Investigate“ Aufgabe aufgeklärt werden, Bodenabwehrstellungen (SAM-Sites) wurden über eine „Suppress“ Aufgabe durch die unbemannten Flugzeuge unterdrückt und bei Radarstellungen sollte keine Aufgabe zugewiesen werden. Für die Delegation mussten die Probanden das jeweilige Objekt antippen, in einem erscheinenden Kontextmenü die zugehörige Aufgabe auswählen und dann in einer Zeitleistendarstellung die Aufgabe an eines der unbemannten Flugzeuge delegieren, indem sie auf einen Zeitpunkt in der Zeitleiste für das jeweilige Flugzeug tippen. Abbildung 37 stellt die vier taktischen Elemente, sowie die Zeitleistendarstellung dar.

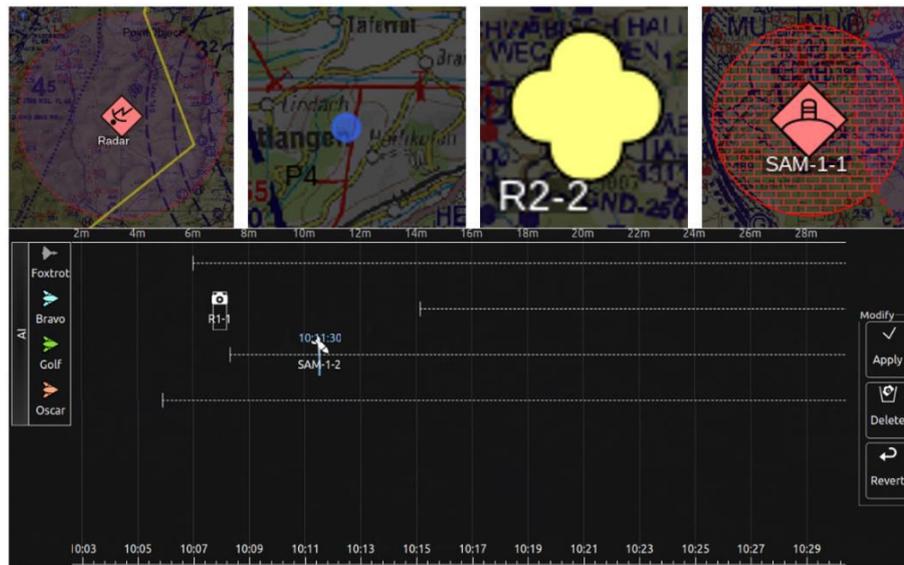


Abbildung 37: Taktische Elemente für das Elektrokardiografbasierte Experiment und Zeitleistendarstellung

Jede Phase mit einer Delegationsaufgabe hatte pro Objekttyp zwei Objektinstanzen in der Karte dargestellt. Somit hatte jede solche Phase acht Objekte dargestellt. Als zusätzliche Anforderung durfte jedes unbemannte Flugzeug nur zwei Aufgaben erhalten und durfte nicht den gleichen Aufgabentyp zweimal erhalten.

Fliegen/Tiefflug

Der dritte Aufgabentyp ist das Fliegen des Kampfflugzeugs entlang einer vorgegebenen Route. Um keine Variation in der Schwierigkeit dieser Aufgabe zu erzeugen, enthielt jede Flugphase zwei Routenabschnitte, die in einem 110 Grad Winkel zueinanderstehen. In den ersten vier Schwierigkeitsstufen mussten die Probanden eine Flughöhe von 5000ft über dem Meeresspiegel, sowie eine Geschwindigkeit von 500kts halten. In der fünften Schwierigkeitsstufe wurde die Höhenvorgabe durch eine Radarhöhenvorgabe von 500ft über dem Boden ausgetauscht, wodurch die Probanden die Höhe kontinuierlich an das Terrain anpassen mussten.

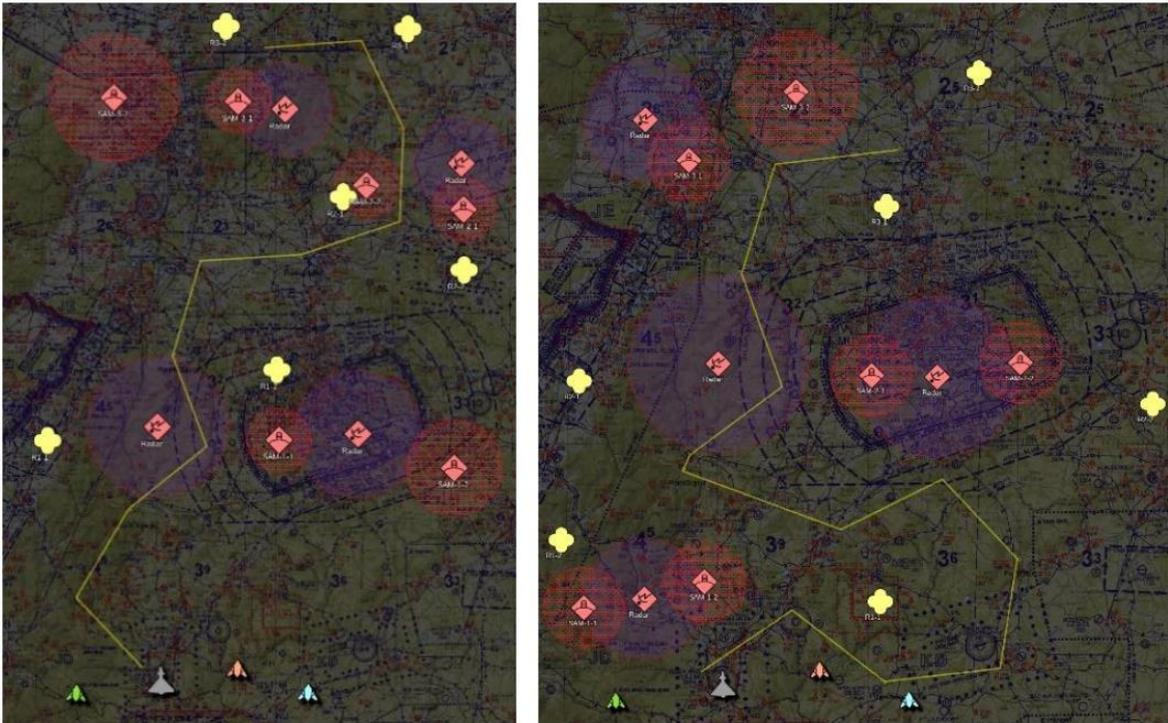


Abbildung 38: Unterschiedliche Missionen für das Elektrokardiografbasierte Experiment

Für das Experiment wurden zwei unterschiedliche Missionen gestaltet (siehe Abbildung 38), um die Reihenfolge der Schwierigkeiten zu variieren, um den Einfluss von Trainingseffekten zu minimieren. Da die Probanden zuvor keine Flugerfahrung aufwiesen wurde dieser Einfluss als hoch eingeschätzt. Die Probanden wurden zufällig in zwei Gruppen eingeteilt und sind jeweils eine der zwei Missionen geflogen.

Bei der Gestaltung der Missionen wurde zusätzlich berücksichtigt, dass das zugrundeliegende Terrain für die Tiefflugkomponente einen Einfluss auf die Schwierigkeit hat. Aus diesem Grund befindet sich der Tiefflugabschnitt in beiden Missionen an demselben Ort. Die Reihenfolge der Schwierigkeiten für die beiden Missionen wird in Tabelle 9 dargestellt.

Tabelle 9: Reihenfolge der Schwierigkeiten der Varianten des Elektrokardiografbasierten Experiments

Phase	Mission 1	Mission 2
1	Schwierigkeit I	Schwierigkeit II
2	Schwierigkeit V	Schwierigkeit III
3	Schwierigkeit III	Schwierigkeit I
4	Schwierigkeit IV	Schwierigkeit V
5	Schwierigkeit II	Schwierigkeit IV

4.3.2.2 Durchführung

Das Experiment wurde mit zehn männlichen Studenten der Universität der Bundeswehr durchgeführt. Alle Versuchsteilnehmer waren zwischen 21 und 24 Jahren alt und hatten keinerlei Flugerfahrung. Jeder der Probanden erhielt eine Einführung in die Simulationsumgebung inklusive einer Erklärung wie das Flugzeug gesteuert wird. Daraufhin durfte jeder Proband sich mit dem Fliegen vertraut machen und die anderen Aufgaben durchführen.

Nach dem Training wurde jedem Probanden die zugehörige Mission mit der Reihenfolge der durchzuführenden Phasen präsentiert und das Experiment gestartet. Während des Experiments wurden die EKG-Daten der Probanden mit einem BioPac Bionomadix Sensor (BIOPAC Systems Inc.) aufgenommen. Am Ende jeder Phase wurde die Simulation gestoppt, um einen NASA-TLX Fragebogen auszufüllen, in dem die Probanden die Beanspruchung der vergangenen Phase bewerten. Daraufhin wurde dem Probanden erneut mitgeteilt welche Aufgaben in der nächsten Phase gefordert sind und die Simulation wurde fortgesetzt. Eine Phase dauerte aufgrund der Geschwindigkeits- und Routenvorgaben ca. drei Minuten.

Nach dem Experiment wurden die Herzdaten mit einem Pan-Tompkins-QRS-Detektionsalgorithmus (Pan und Tompkins 1985) verarbeitet und die resultierenden RR-Intervalle (die Zeitabstände zwischen zwei Herzschlägen) mit der Software Kubios-HRV (KUBIOS OY.) analysiert. Bei der Analyse wurde lediglich das Zeitfenster von 30 Sekunden nach Beginn der Phase bis zu der zwei Minuten Marke untersucht. Die ersten 30 Sekunden wurden nicht gewertet, um den Probanden Zeit zu geben sich nach dem Anhalten der Simulation wieder hineinzusetzen und bestehende Abweichungen in der Flughöhe zu korrigieren. Die letzte Minute wurde nicht bewertet, da die Probanden zu dieser Zeit häufig die Delegationsaufgabe bereits abgeschlossen haben und somit diese Zusatzbelastung

weggefallen ist. Des Weiteren wurden keine manuellen Korrekturen an der Erkennung durchgeführt, da diese im Falle eines Einsatzes in einem adaptiven Automationssystem ebenfalls nicht möglich wären.

Neben den Herzdaten wurde auch die Leistung der Probanden bewertet, indem die quadratische Abweichung zwischen der tatsächlichen- und der vorgegebenen Flughöhe berechnet wurde. Außerdem wurden die Anzahl der richtig beantworteten Kopfrechenaufgaben sowie der korrekt delegierten Aufgaben an die unbemannten Flugzeuge bewertet.

4.3.2.3 Ergebnisse

Die Befragung der Probanden durch den NASA-TLX Fragebogen hat ergeben, dass zwischen den einzelnen Phasen signifikante Beanspruchungsunterschiede lagen ($t(10) = -5.1, p < 0.001$ für I und II; $t(10) = -2.3, p = 0.048$ für II und III; $t(10) = -4.5, p = 0.001$ für III und IV; $t(10) = -2.5, p = 0.03$ für IV und V). Abbildung 39 stellt die subjektiven Beanspruchungswerte der fünf Phasen in einem Boxplot gegenüber.

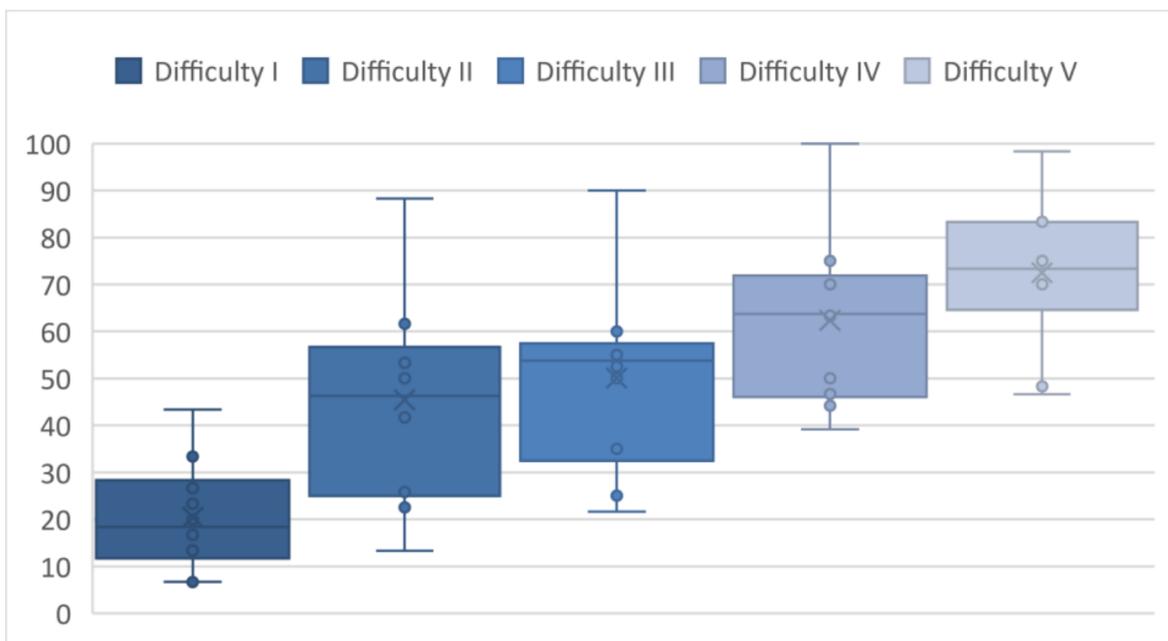


Abbildung 39: NASA-TLX Bewertungen der fünf Phasen des Elektrokardiograph-basierten Experiments

Die Analyse der Flugleistung, gemessen an der Abweichung von der Höhenvorgabe (dargestellt in Abbildung 40), zeigt einen Anstieg des Fehlers innerhalb der ersten vier Schwierigkeitsphasen. Gemessen anhand des Medians gibt es hierbei einen Anstieg von der ersten auf die zweite Schwierigkeit. Zwischen der zweiten und der dritten Schwierigkeitsstufe bleibt die Höhenabweichung ungefähr gleich und Stufe vier zeigt einen

starken Anstieg im Vergleich zu den vorherigen beiden Stufen. Bei der vermeintlich schwierigsten Phase zeigt sich allerdings eine verbesserte Leistung in der Flugaufgabe als in den Phasen zwei, drei und vier.

Bei der Leistung in der „Kopfrechnen“ Aufgabe zeigte sich eine höhere Leistung in der Schwierigkeitsstufe III (Median: 18.5, MAD: 5.5), als in den beiden Phasen, die zusätzlich die Delegationsaufgabe beinhalteten (Schwierigkeit IV Median: 10, MAD: 4 und Schwierigkeit V Median: 11.5, MAD: 4.5). Bei der Delegationsaufgabe traten nur sehr vereinzelt Fehler auf. In den Schwierigkeitsstufen II und IV ist bei den insgesamt 48 notwendigen Delegationen nur jeweils eine nicht durchgeführte Zuweisung aufgetreten und in Schwierigkeitsstufe V ist diese Anzahl auf vier gestiegen.

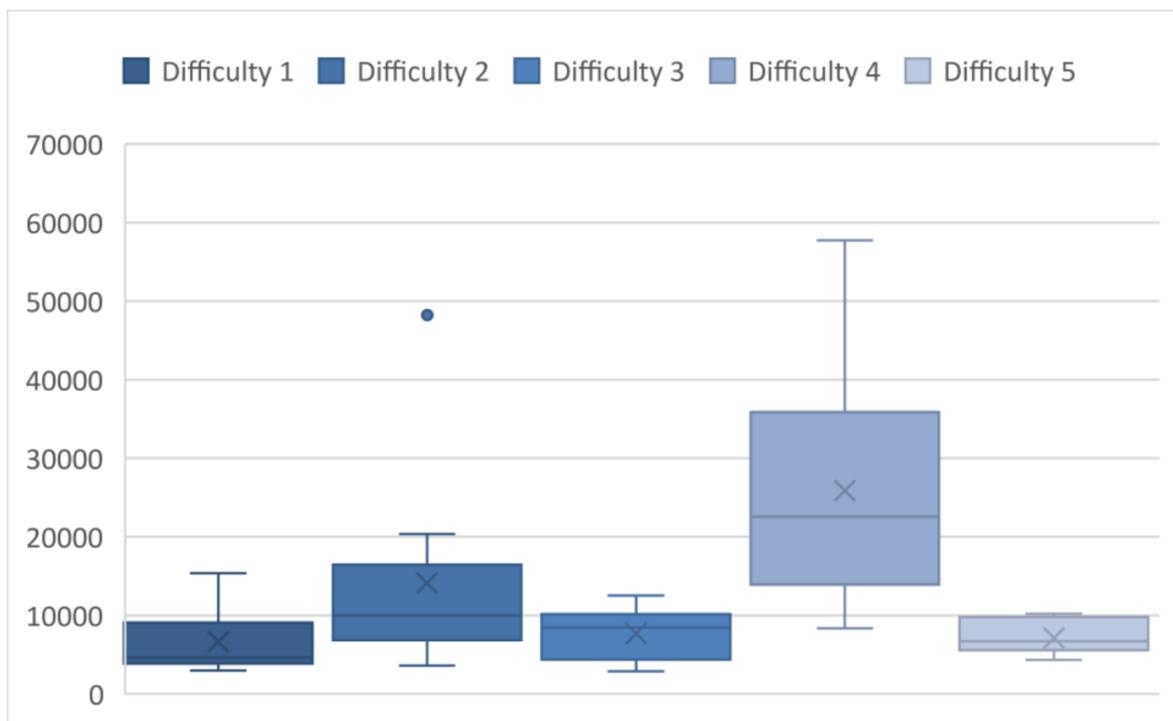


Abbildung 40: Leistung in der Flugaufgabe des Elektrokardiografbasierten Experiments in Form der kumulierten Abweichung von der Höhenvorgabe in ft

Bei der Analyse der Herzrattendaten wurde festgestellt, dass ein Datensatz aufgrund von Fehlern bei der Aufnahme leider nicht verwendet werden konnte. Die verbliebenen Daten wurden wie beschrieben mit einer Pan-Tompkins QRS-Detektion verarbeitet. Hierbei wurde festgestellt, dass aufgrund sehr hoher T-Wellen eines Versuchsteilnehmers viele falschpositive Schläge erkannt wurden und deshalb ein selbstentwickelter Peak-Detektor stattdessen eingesetzt wurde. Des Weiteren wurden insgesamt zwei Phasen aus der Analyse ausgeschlossen (Teilnehmer 6 Schwierigkeit II und Teilnehmer 10 Schwierigkeit I), weil deren Daten zu stark verrauscht waren. Für die Vergleichbarkeit der Teilnehmer

untereinander wurden die resultierenden Daten weiterhin anhand der individuellen Minima und Maxima der einzelnen Versuchsteilnehmer auf einen Wert zwischen 0 und 1 normiert.

Die somit gewonnenen Daten wurden daraufhin mit den Antworten aus den NASA-TLX Fragebögen korreliert. Hierbei wurden die Beanspruchungen der subjektiven Fragebögen als Grundwahrheiten zugrunde gelegt und die Korrelationsstärke der betrachteten Herzparameter als Bewertung des Herzparameters genutzt. Tabelle 10 zeigt die Korrelationen der Herzraten sowohl individuell für jeden Probanden als auch den Durchschnitt, für die evaluierten Parameter. Die Parameter sind die bereits in Tabelle 3 beschriebenen Parameter. Hierbei treten bei vereinzelt Probanden nahezu perfekte Korrelationen zwischen beispielsweise den durchschnittlichen Herzraten und der subjektiven Beanspruchungsbewertungen, aber für den gleichen Parameter auch entgegengerichtete Korrelationen auf. Dadurch ergibt sich im Mittel immer noch eine verwertbare Korrelation.

Tabelle 10: Korrelation der Herzparameter mit den durchschnittlichen NASA-TLX Bewertungen im elektrokardiografbasierten Experiment

Participant	Avg. HR	RMSSD	RR tri	HF	SD2
2	0.98	0.93	0.75	0.02	0.74
3	0.72	-0.25	0.30	-0.93	0.72
4	-0.19	-0.15	0.96	-0.86	0.94
5	0.98	0.63	0.64	0.44	0.08
6	-0.37	0.58	-0.56	0.16	0.62
7	0.96	-0.16	0.42	-0.88	0.59
8	0.49	0.67	0.24	-0.17	0.82
9	0.72	0.33	0.52	-0.33	0.32
10	0.42	-0.64	0.57	-0.15	-0.25
Avg	0.52	0.22	0.42	-0.30	0.51

Für die Herzparameter mit den stärksten Korrelationen (Herzrate und SD2) wurden weiterhin abhängige t-Tests ausgewertet, um festzustellen welche Veränderungen zwischen den Schwierigkeitsphasen signifikant sind. Bei der Herzrate (dargestellt in Abbildung 41) konnte hierbei keine Signifikanz von zwei aufeinander folgenden Phasen festgestellt werden. Allerdings war der Unterschied zwischen Schwierigkeit II und sowohl Schwierigkeit IV ($t(9) = -2.65$, $p = 0.03$) als auch Schwierigkeit V ($t(9) = -2.93$, $p = 0.02$) signifikant.

Gleichermaßen gab es einen signifikanten Unterschied von Schwierigkeit I zu den Schwierigkeiten IV ($t(9) = -4.60, p = 0.001$) und V ($t(9) = -2.71, p = 0.03$).

Die Analyse der SD2 Werte der Poincaré Abbildung des RR Abstands zeigt einen signifikanten Anstieg von Schwierigkeit I zu Schwierigkeit III ($t(9) = -2.89, p = 0.02$) und Schwierigkeit IV ($t(9) = -2.81, p = 0.02$). Zusätzlich konnte ein signifikanter Unterschied zwischen Schwierigkeit II und Schwierigkeit IV ($t(9) = -3.49, p = 0.008$) festgestellt werden. Abbildung 42 stellt die zugrundeliegenden Daten grafisch dar.

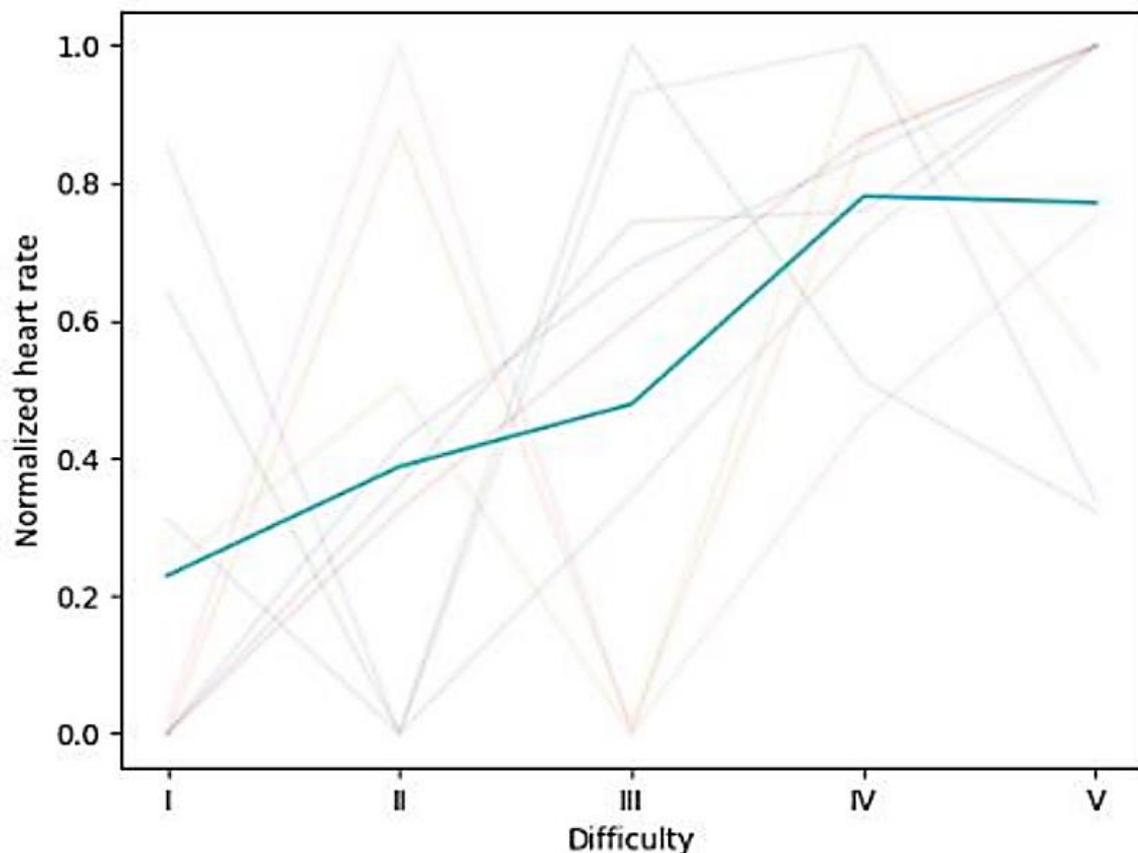


Abbildung 41: Normalisierte Herzrate zu den Schwierigkeitsstufen im elektrokardiografbasierten Experiment. Die türkise Linie zeigt den Mittelwert aller Teilnehmer, während die semitransparenten Linien jeweils einen Teilnehmer darstellen.

4.3.2.4 Diskussion der Ergebnisse

Die subjektive Bewertung der Beanspruchung durch den NASA-TLX zeigt ein erfolgreiches Experimentaldesign insofern, dass mehrere eindeutig voneinander differenzierbare Schwierigkeitsstufen erzeugt werden konnten. Die sinkende Leistung der Versuchsteilnehmer mit steigender Versuchsschwierigkeit in den Schwierigkeiten I bis IV zeigt dabei, dass die induzierte Beanspruchung hoch genug war, um Leistungseinbrüche zu erzeugen. Die erhöhte Leistung in der Flugaufgabe in der Schwierigkeit V könnte sich durch

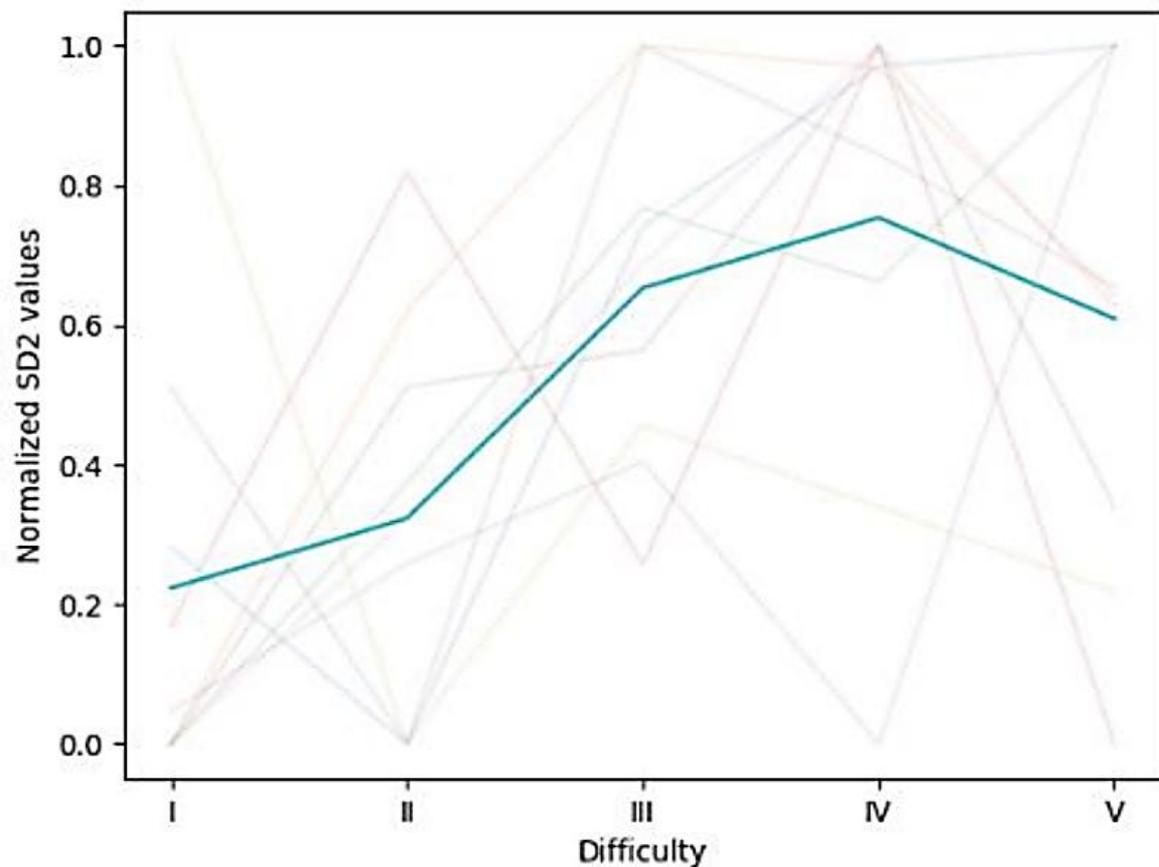


Abbildung 42: Normalisierte SD2 Werte zu den Schwierigkeitsstufen im elektrokardiografbasierten Experiment. Die türkise Linie zeigt den Mittelwert aller Teilnehmer, während die semitransparenten Linien jeweils einen Teilnehmer darstellen

das verbesserte visuelle Feedback begründen. In der Tiefflugaufgabe mussten die Versuchsteilnehmer die Höhe relativ zum Terrain anpassen und hatten somit eine visuelle Orientierungshilfe. Hingegen bei der Flugaufgabe mit einer konstanten Höhenvorgabe mussten die Teilnehmer sich an einer Zahlendarstellung orientieren und hatten somit keine visuelle Rückmeldung in der peripheren Sicht. Dies führt dazu, dass trotz der erhöhten Schwierigkeit der Tiefflugaufgabe (die Anpassung an das Terrain erfordert regelmäßige Korrekturen) eine verbesserte Leistung erzielt werden konnte. Gleichzeitig kann dies auch eine Begründung für die erhöhte subjektive Beanspruchung in Schwierigkeit V darstellen, obwohl die Leistung und die Herzparameter eine geringere Beanspruchung in dieser Phase signalisieren.

Die Ermittlung der Beanspruchung auf Basis der herzbasierten Parameter zeigt im Mittel eine Korrelation mit der subjektiven Bewertung der Beanspruchung, allerdings treten in den Einzelverläufen der Versuchsteilnehmer stark variierende Verläufe auf. Lediglich die Herzrate kann als zuverlässiger Indikator für eine erhöhte Beanspruchung bewertet werden. Weiterhin zeigen sich individuelle Differenzen zwischen den Versuchsteilnehmern in den

Absolutwerten der herzbasierten Parameter. Diesen Effekten wurde in diesem Experiment entgegengewirkt, indem die Daten pro Proband mithilfe des Minimums und des Maximums des jeweiligen Parameters normalisiert wurden. Auch in der Literatur werden aus diesem Grund häufig Vergleiche zu einer „Baseline-Messung“ gezogen, um Aussagen über die Auswirkung von mentaler Beanspruchung zu beurteilen. Bei einer solchen Baseline handelt es sich um eine Messung, bei der die Versuchsteilnehmer gebeten werden, sich zu entspannen um einen Ruhezustand als Referenzmesswert zu generieren. Allerdings variiert zwischen mehreren Versuchsteilnehmern nicht ausschließlich der Ruhezustand, sondern ebenfalls die Effektstärke der beanspruchungsinduzierten Veränderung des jeweiligen psychophysiologischen Parameters. Diese Effektstärken können bis dato nicht über eine Referenzmessung erfasst werden und erzeugen dadurch weiterhin interindividuelle Unterschiede.

4.3.2.5 Bewertung der Messmethode

Es konnte gezeigt werden, dass mithilfe von elektrokardiografbasierten Messungen Unterschiede aufgrund von Beanspruchung festgestellt werden können. Innerhalb des Experiments konnten erfolgreich mehrere Stufen von Beanspruchung voneinander abgegrenzt werden und die Messmethode kann trotz des verkürzten Zeitfensters relevante Aussagen generieren.

Somit hat das Experiment ergeben, dass insbesondere die Herzrate ein Indikator für Beanspruchung ist. Allerdings variieren Absolutwerte der Probanden sehr stark, wodurch allgemeingültige Aussagen schwierig sind. Eine mögliche Lösung dieses Problem anzugehen ist die Betrachtung von relativen Änderungen der Herzrate als Indikator für einen starken Beanspruchungsanstieg oder -abfall.

Robustheit

Die Messung der Herzratenparameter kann dank automatisierten Korrekturfunktionen, wie sie beispielsweise in Kubios HRV implementiert sind, sehr robust aus dem EKG-Signal extrahiert werden. Zusätzlich konnte gezeigt werden, dass eine Änderung der kognitiven Beanspruchung zu mehreren signifikant voneinander unterscheidbaren Stufen der einzelnen Herzparameter führen. Andererseits kann eine Änderung der Herzrate ebenfalls durch physische Beanspruchung entstehen. Kann diese physische Beanspruchung beim Einsatz des adaptiven Systems ausgeschlossen werden, wie in der verwendeten Simulationsumgebung ist die Robustheit **sehr hoch**.

Sensitivität

Durch die starke Variation in der Beanspruchung innerhalb des Experiments konnten mehrere Stufen von Beanspruchung unterschieden werden. Allerdings konnte keine klare

Zuordnung zwischen einer erkennbaren Beanspruchungsstufe und einer Überbeanspruchung identifiziert werden. Außerdem stellt eine dreistufige Beanspruchungsklassifikation im Vergleich zu modellbasierten Ansätzen oder Fragebogenbewertungen eine niedrige Unterscheidungsfähigkeit dar. Dadurch ist die Sensitivität insgesamt als **mittelmäßig** zu bewerten.

Diagnostizität

Laut der Literatur sind herzbasierte Messverfahren für mentale Beanspruchungsbewertung auf die Erkennung kognitiver Beanspruchung begrenzt. Somit kann keine Unterscheidung zwischen Beanspruchungsdimension durchgeführt werden. Aufgrund dessen wurde innerhalb des Experiments vorwiegend kognitive Beanspruchung durch die Nebenaufgaben stimuliert. Insgesamt ist die Diagnostizität von Herzparametern damit als **niedrig** einzustufen.

Individualität

Die Herzparameter sind in der Lage individuelle Beanspruchungssituationen aufzudecken. Um dieser Individualität gerecht zu werden, wird häufig ein „*Baseline*“-Messung durchgeführt, bei der die Herzparameter der Menschen in einer Ruhesituation aufgenommen werden. Die Schwierigkeit der Messmethode liegt in der individuellen Unterscheidung zwischen Beanspruchung und Überbeanspruchung, da zusätzlich zu der *Baseline* auch die Veränderung der Herzparameter aufgrund von Beanspruchung sich individuell ändert. Insgesamt ist die Individualität von Herzparametern **hoch**.

Generalisierbarkeit

Die Bewertung der Beanspruchung durch elektrokardiografische Messungen benötigt keinerlei Kontextinformationen und ist für mehrere unterschiedliche Messverfahren einsetzbar. Die Herzrate im speziellen kann hierbei, neben der Verwendung eines Elektrokardiografen, zusätzlich kamerabasiert durch Veränderungen der Hautfarbe (Fernandes et al. 2020) oder mit Hilfe von beispielsweise Armbanduhren durch Veränderungen des Blutvolumens ermittelt werden. Diese Vielseitigkeit sorgt dafür, dass die Generalisierbarkeit des Messverfahrens **sehr hoch** ist.

Direkte Anwendbarkeit

Durch die Begrenzung auf kognitive Beanspruchung und die Möglichkeit das Messverfahren einzusetzen, ohne etwas über die Einsatzdomäne wissen zu müssen, kann keinerlei Aussage über die Ursache der Beanspruchung generiert werden. Zusätzlich konnte innerhalb des Experiments keine eindeutige Identifikation Überbeanspruchung gegenüber einer hohen Beanspruchung durchgeführt werden. Somit ist die herzparameterbasierte Ermittlung von Beanspruchung ausschließlich dazu geeignet, um Beanspruchung festzustellen, nicht aber

dazu Möglichkeiten zur Adaptivität zu identifizieren. Somit ist die direkte Anwendbarkeit des Verfahrens **sehr niedrig**.

4.4 Kontextbehaftete Ermittlungsmethoden

Im Rahmen der kontextbehafteten Ermittlungsmethoden wurde zum einen die Möglichkeit zur Erkennung von selbstadaptiven Strategien untersucht und ermittelt, ob diese Erkennung eine individuelle Beanspruchungsermittlung erzeugen kann. Die Methode ist dabei auf die Beanspruchungsermittlung in einer wiederkehrenden Aufgabensituation fokussiert. Demgegenüber steht ein Experiment über die Erkennung der Aktivität des Menschen. Mithilfe der Aktivität können die Beanspruchungsmodelle aus Kapitel 3 verwendet werden, um eine Beanspruchungsschätzung auf Basis der Aktionen zu generieren. Zusätzlich erlaubt die erkannte Aktivität eine verbesserte Beschreibung des Aufgabenkontexts, in dem eine Problemsituation aufgetreten ist, wodurch eine Automation besser an den Menschen angepasst werden kann

4.4.1 Verhaltensbasiertes Experiment

Ziel des verhaltensbasierten Experiments war es Verhaltensänderungen zu identifizieren, die auftreten, wenn eine erhöhte Beanspruchungssituation eintritt. Hierbei soll insbesondere die Möglichkeit zur Erkennung allgemeiner Verhaltensänderungen untersucht werden, wodurch solche Verhaltensmuster auch in neuen Anwendungsdomänen ermittelt werden können.

4.4.1.1 Experimentaldesign

Beanspruchungsinduzierte Verhaltensänderungen können nur erkannt werden, wenn die gleiche Aufgabensituation mehrmals eintritt und dadurch ein Modell des normativen Verhaltens erstellt werden kann. Dies erlaubt eine Erkennung einer Abweichung von diesem normativen Verhalten.

Aus diesem Grund war es für dieses Experiment notwendig eine gleichbleibende Aufgabe sehr häufig durchzuführen und dabei unterschiedliche Belastungslevel zu erzeugen.

4.4.1.2 Beschreibung

Das Experiment bestand aus zwei Aufgaben die gleichzeitig durchgeführt mussten. Die erste Aufgabe war es eine vorgegebene Route im Tiefflug abzufliegen. Die zweite Aufgabe war die Klassifikation von regelmäßig eingespielten Bilddaten. Für die Tiefflugaufgabe gab es eine Höhenvorgabe von 100 bis 300 ft über dem Grund die bestmöglich eingehalten werden

sollte. Die Klassifikationsaufgabe enthielt reale Aufklärungsbilder, bei denen der Proband die Art des abgebildeten Objekts (z.B. Vehikel) und die militärische Zugehörigkeit zuordnen musste. Abbildung 43 stellt exemplarisch zwei der eingespielten Klassifikationsbilder dar. Das linke Bild musste hierbei mit der Zugehörigkeit „befreundet“ und mit der Art „Gebäude“ klassifiziert werden, wobei das rechte Bild ein „Vehikel“ mit der Zugehörigkeit „befeindet“ darstellt.

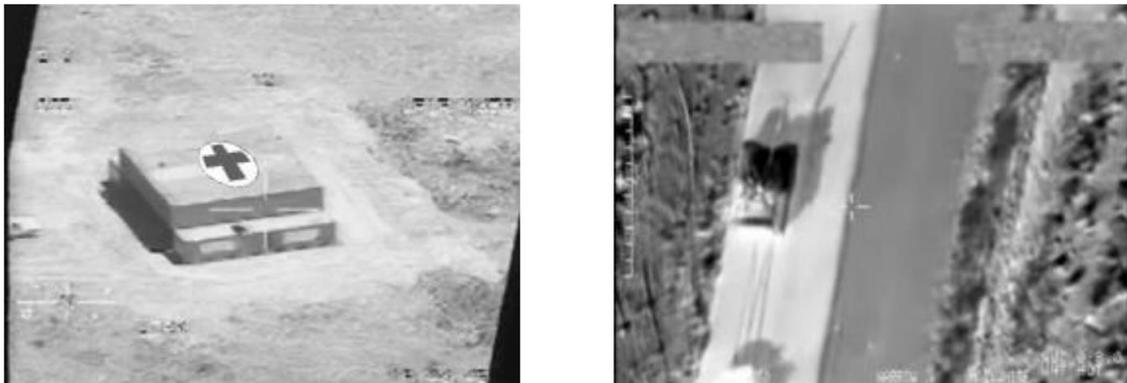


Abbildung 43: Eingespielte Klassifikationsbilder im Verhaltensexperiment

Zusätzlich mussten in jeder Testmission vier Angriffe auf vordefinierte Ziele durchgeführt werden, wobei der Angriff in einem zehn Sekunden Zeitfenster stattfinden musste. Die Belastung des Probanden wurde durch die Häufigkeit der eingespielten Aufklärungsbilder moduliert. Hierbei wurde jede der vier geflogenen Testmissionen in fünf Missionssegmente aufgeteilt, wobei die ersten vier Phasen jeweils mit einem Angriff abgeschlossen wurden. Nach jedem der Angriffe wurde die Simulation pausiert, damit der Proband durch einen NASA-TLX Fragebogen die vergangene Beanspruchungssituation bewerten kann.

Die Probanden in diesem Experiment waren sechs Luftwaffenpiloten in einem Alter von 29 bis 56 Jahren. Das Durchschnittsalter der Piloten lag bei 42 Jahren.

4.4.1.3 Ergebnisse

In diesem Abschnitt werden die Erkenntnisse aus unseren Experimenten mit Piloten der deutschen Luftwaffe beschrieben. Zunächst werden die Ergebnisse der subjektiven Beanspruchungsbeurteilung dargestellt. Anschließend werden die Leistung und die beobachteten Verhaltensanpassungen (selbstadaptive Strategien) untersucht.

Die Leistungsmaße sowohl für die primäre als auch für die sekundäre Aufgabe zeigen keine eindeutige Korrelation mit der Dichte der Bilder, die vom Piloten klassifiziert werden mussten. Dies deutet darauf hin, dass die Bilddichte als alleiniger Treiber für die Schwierigkeit eines Missionssegments unzureichend ist. Wir haben die Struktur des

angetroffenen Geländes, approximiert durch die Standardabweichung der Geländehöhe, als zweiten wichtigen Einflussfaktor für die Schwierigkeit eines Missionssegments identifiziert. Auf der Grundlage dieser Erkenntnisse wurden die Missionssegmente in drei Cluster mit ähnlicher Geländeschwierigkeit und zunehmender Bilddichte (niedrig, mittel, hoch) und einen Cluster mit mittlerer Bilddichte und hoher Geländeschwierigkeit eingeteilt (siehe Abbildung 44). Mit diesen Clustern ist es möglich, die Auswirkungen einer erhöhten Bilddichte sowie einer erhöhten Geländeschwierigkeit auf die erfasste Arbeitslast und die Leistung unabhängig voneinander zu bewerten.

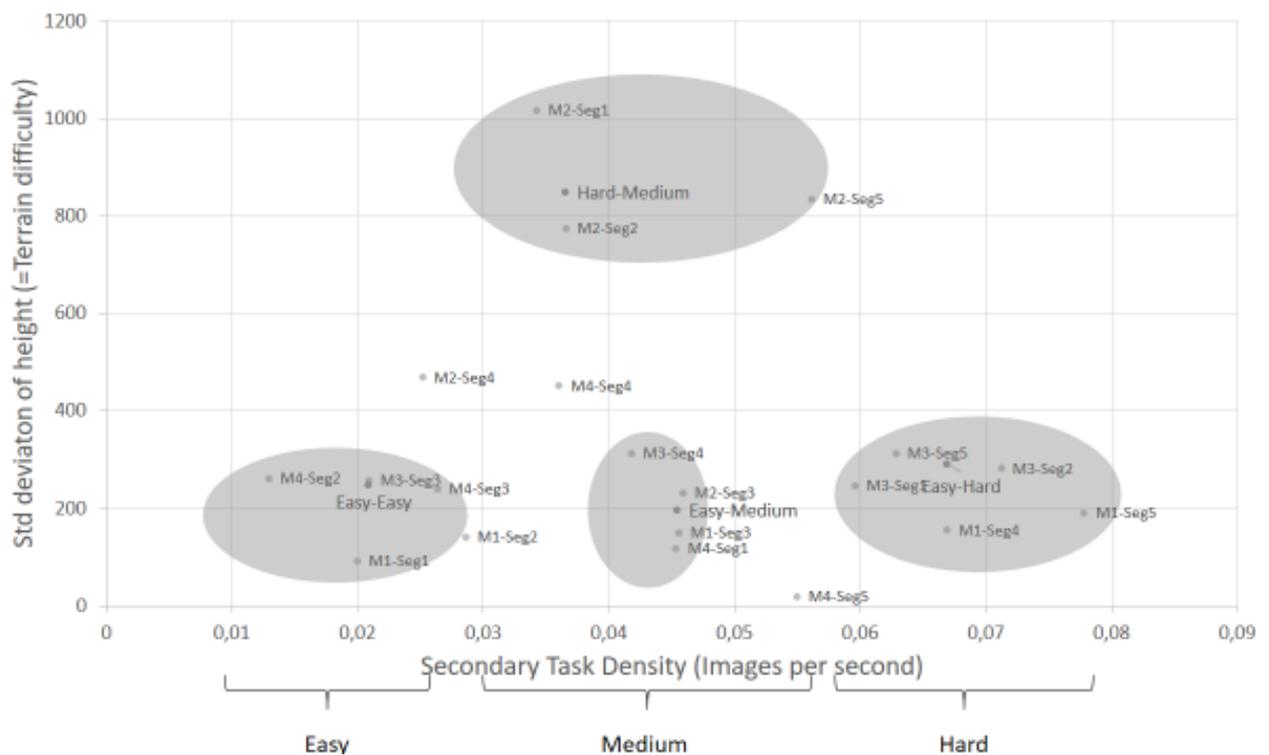


Abbildung 44: Einteilung in vier Schwierigkeitscluster. MX-SegY ist eine kurzschreibweise für Mission X Segment Y

Subjektive Beanspruchungsbewertung

Die kombinierten Arbeitsbelastungsbewertungen aus den NASA-TLX Fragebögen in Abbildung 45 zeigen eine mittlere bis hohe Arbeitsbelastung in allen Experimenten. Dies ist ein guter Hinweis darauf, dass der Versuchsaufbau die Arbeitsbelastung wie beabsichtigt anregte. Darüber hinaus beobachteten wir einen unbedeutenden Anstieg der Arbeitsbelastung mit zunehmender Bilddichte (leicht-leicht->leicht-mittel->leicht-schwer). Zwischen den Clustern mit unterschiedlichem Schwierigkeitsgrad des Geländes (Easy-Medium vs. Hard-Medium) gibt es hingegen keine Veränderung in den subjektiven Belastungsbewertungen. Sowohl der Mittelwert als auch der Median dieser Bewertungen sind bei beiden Clustern ähnlich. Abbildung 45 zeigt Boxplots der

Arbeitsbelastungsbewertungen, die den vier verschiedenen Clustern zugeordnet wurden. Die Whisker stellen die Extremwerte (Minimum, Maximum) der jeweiligen Bewertungen dar.

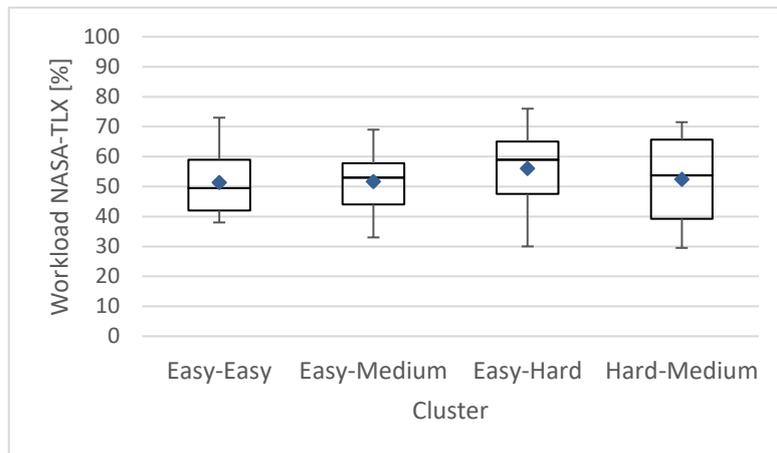


Abbildung 45: NASA-TLX Bewertungen der vier Cluster

Leistungsbewertung

Allen Probanden gelang es, die Missionsziele anzugreifen, mit Ausnahme eines Probanden, der sein erstes Ziel in der ersten Mission verfehlte. Der eine fehlgeschlagene Zielangriff wird einem Mangel an Training zugeschrieben. Was die vorgegebenen Zeitvorgaben betrifft, so gab es im Allgemeinen nur sehr wenige Abweichungen von den geforderten Time-Over-Target-Intervallen. Eine Ausnahme bildete das zweite Segment der Mission drei, in dem bis auf eine Person alle Probanden das Zeitintervall nicht einhielten. Das dritte Leistungskriterium für die Primäraufgabe, nämlich die Einhaltung des Höhenbereichs, zeigt eine Korrelation mit der Dichte der Sekundäraufgabe. Eine höhere Dichte dieser Klassifikationsaufgabe führt zu einer vermehrten Verletzung des Höhenbereichs, wie die Mittelwerte für die Cluster Easy-Easy, Easy-Medium und Easy-Hard in Abbildung 46 zeigen.

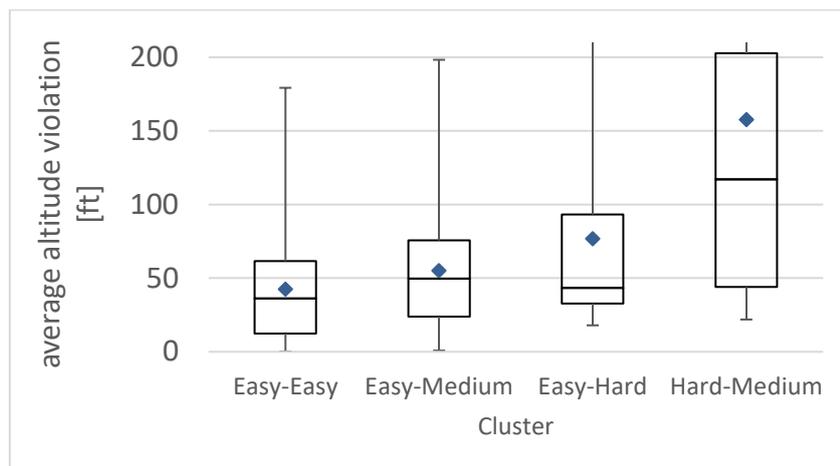


Abbildung 46: Durchschnittliche Verletzung der Höhenvorgabe innerhalb der Cluster

Bei der sekundären Aufgabe waren alle Probanden in der Lage, die 56 Bilder pro Mission zu verarbeiten, die von den UAVs geliefert wurden. Die durchschnittliche Bearbeitungszeit für eine Identifikationsaufgabe wurde weder durch die Schwierigkeit des Geländes noch durch die Bilddichte beeinflusst, wie in Tabelle 11 dargestellt. Im Gegensatz dazu stieg die Zeit zwischen dem Erscheinen einer Klassifizierungsaufgabe und dem Beginn ihrer Bearbeitung mit zunehmender Aufgabenhäufigkeit signifikant an, und zwar von einem Mittelwert von 7,5 Sekunden (Stichprobenumfang = 180; std. Abweichung = 6,3s) im Easy-Easy-Cluster auf 9,41 Sekunden (Stichprobenumfang = 450; std. Abweichung = 8,4s) im Easy-Hard-Cluster. Wie schon bei der Bearbeitungszeit zeigt sich auch bei diesem Leistungsmaß kein Einfluss der Geländeschwierigkeit.

Tabelle 11: Mittelwert und Standardabweichung der Zeit für die Sekundäraufgabe

Sektion	Easy-Easy	Easy-Medium	Easy-Hard	Hard-Medium
Mittelwert	4.53	4.23	3.91	4.19
Standardabweichung	2.05	1.34	0.94	1.07

Wie in (Mund et al. 2017) beschrieben, konnten bei der Auswertung des Experiments die folgenden vier unterschiedliche Verhaltensänderungen festgestellt werden:

- Lockerung der Höhenvorgabe
- Verzögerung der Aufklärungsaufgabe
- Umpriorisierung
- Batch-Verarbeitung

Die Nutzung dieser Strategien ist vorwiegend in den Missionsphasen mit einer höheren Belastung aufgetreten was darauf schließen lässt, dass sie eine Form der in (Sperandio 1971)

beschriebenen „Selbstadaptiven Strategien“ darstellen. Im Folgenden werden diese Strategien genauer betrachtet.

Lockerung der Höhengvorgabe

Diese Strategie stellt eine Vereinfachung der Primäraufgabe dar. Die Probanden haben bewusst die Höhengvorgabe verletzt, um mentale Ressourcen freizulegen. Diese Lockerung trat vorwiegend in den Missionssegmenten mit einer erhöhten Bilddichte auf. Das Nutzen dieser Strategie wurde im Rahmen eines Debriefings von den Probanden bestätigt.

Verzögern der Aufklärungsaufgabe

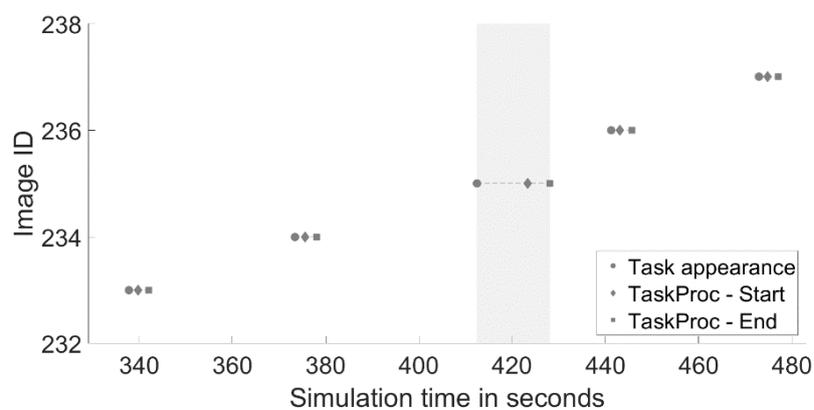


Abbildung 47: Visualisierung der Verzögerung der Aufklärungsaufgabe

Bei dieser Strategie wurde das Bearbeiten der Sekundäraufgabe aufgrund der Beanspruchung durch die Primäraufgabe verzögert. Wie in Abbildung 47 dargestellt wurde die Sekundäraufgabe bei dem Anwenden der Verzögerungsstrategie erst nach ca. 10 Sekunden begonnen (grau markierter Bereich), während sie bei regulärer Aufgabendurchführung bereits nach ca. zwei Sekunden bearbeitet wird.

Umpriorisierung

Die Umpriorisierung ist dadurch charakterisiert, dass der Operator beginnt die Bildklassifizierungsaufgabe durchzuführen, indem er durch eine Interaktion mit dem Touchscreen das Bild öffnet, dann aber die Aufgabenbearbeitung unterbricht um sich erneut der Flugaufgabe zu widmen. Abbildung 48 visualisiert die Anwendung der Strategie (rechts) gegenüber regulärer Aufgabendurchführung.

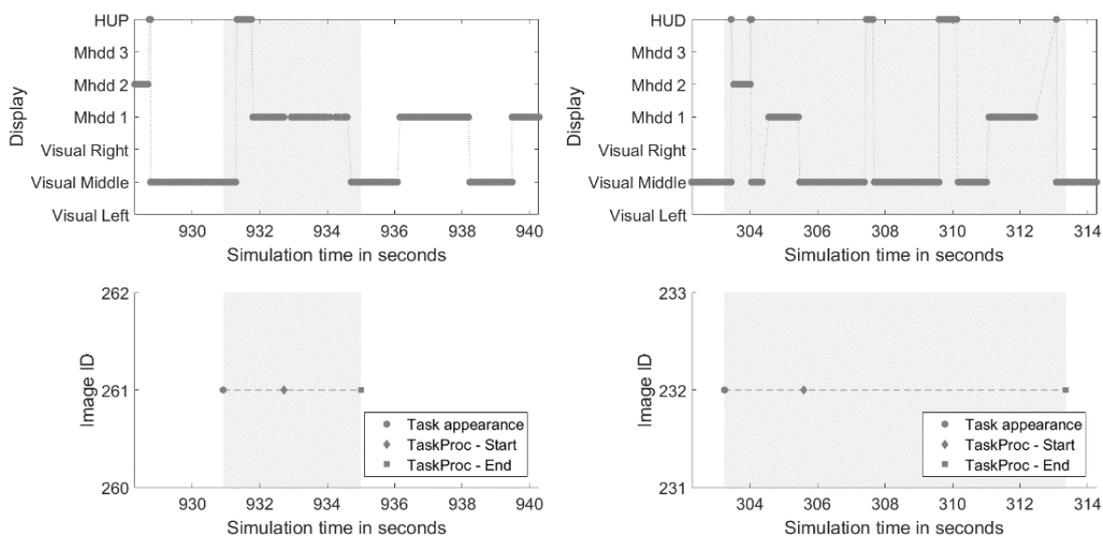


Abbildung 48: Vergleich von Aufgabenabarbeitung mit(rechts) und ohne(links) Anwendung der Strategie "Umpriorisierung"

Bei der regulären Durchführung erscheint der Hinweiston für ein neues Bild und kurz darauf richtet der Operateur seinen Blick auf den Bildschirm „MHDD 1“ auf dem die Klassifikationsaufgabe stattfindet und der Blick verweilt dort, bis die Aufgabe beendet ist. Im Gegensatz dazu wird bei der Durchführung unter Zuhilfenahme der Umpriorisierungsstrategie ebenfalls das Bild geöffnet und der Operateur sieht ebenfalls auf den Bildschirm „MHDD 1“, allerdings wird hierbei die Aufgabendurchführung unterbrochen und der Blick wandert zurück auf die Außensichtsimulation („gekennzeichnet mit „Visual Middle““) um den Fokus auf die Flugaufgabe zu legen. Später wandert der Blick erneut zu der Klassifikationsaufgabe und diese wird unter Inkaufnahme der Verzögerung bearbeitet.

Batch-Verarbeitung

Die letzte der Ermittelten Strategien ist eine Kombination der zwei Strategien *Lockerung der Höhenvorgabe* und *Verzögern der Aufklärungsaufgabe*. Hierbei wird die Bearbeitung der Sekundäraufgabe verzögert, bis sich mehrere Bilder angesammelt haben. In einem Folgeschritt wird die Auflage der Flugaufgabe verletzt, um die angesammelten Bilder direkt nacheinander zu bearbeiten. Eine Befragung der Probanden hat ergeben, dass die beiden Schritte der Strategie bewusst angewendet wurden, um die Aufgabe bewältigen zu können.

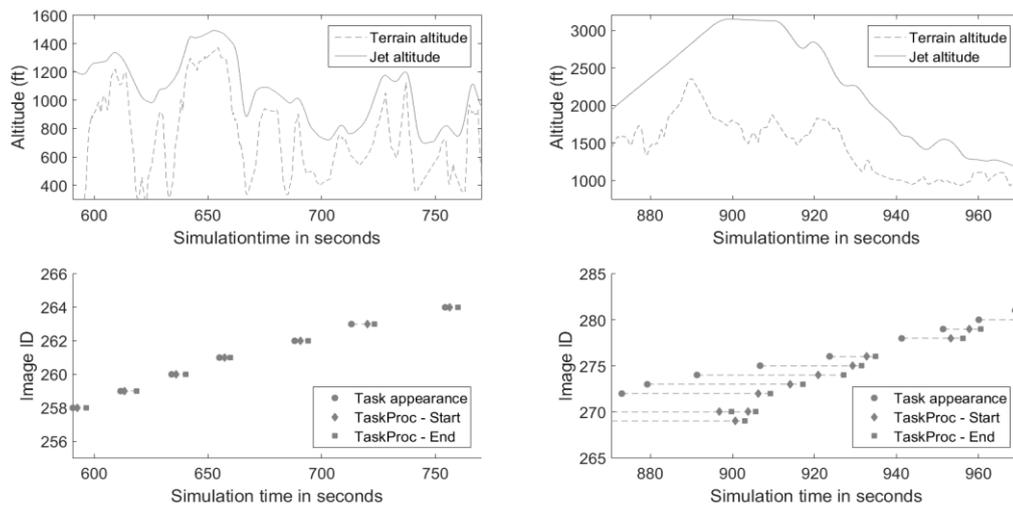


Abbildung 49: Unterschied zwischen regulärer Aufgabendurchführung(links) und Batch-Verarbeitung(rechts)

Abbildung 49 stellt die unterschiedlichen Arten der Aufgabendurchführung dar. Auf der linken Seite wird die Höhenvorgabe konsequent eingehalten und die Klassifikationsaufgabe wird wenige Sekunden nach dem Erscheinen eines neuen Bilds bearbeitet. Im Gegensatz dazu sind die Bilder auf der rechten Seite mehr als 30 Sekunden alt und der Pilot beginnt den Geländefolgeflug zu unterbrechen, um die angesammelten Bilder zu klassifizieren. Hierbei werden die Bilder wie bei der regulären Bearbeitung der Klassifikationsaufgabe innerhalb von ein bis zwei Sekunden nach dem Öffnen des Bildes klassifiziert. Am Ende des Abschnitts geht der Pilot zurück in den Geländefolgeflug, da die angesammelten Bilder fertig klassifiziert sind und auch die Änderung des Geländes hat abgenommen, wodurch die Gesamtbeanspruchung sinkt.

4.4.1.4 Bewertung von Verhaltensbasierten Messungen

Das Experiment hat gezeigt, dass in Phasen erhöhter Belastung Verhaltensänderungen detektiert werden können und somit das Auftreten dieser Verhaltensänderungen ein Indikator für hohe Beanspruchung darstellt. Allerdings konnte nicht jede Verhaltensänderung bei jedem Probanden festgestellt werden. Hierfür gibt es zwei mögliche Ursachen:

- Nicht alle Probanden waren stark genug beansprucht um alle Strategien zu Nutzen
- Welche Strategien genutzt werden und in welchem Maße hängt vom Individuum ab

Welche dieser Ursachen die beschriebenen Unterschiede besser erklärt kann sehr stark von dem gegebenen Aufgabenkontext, sowie dem Ausbildungsstand der Probanden mit dem Mensch-Maschine System abhängen und Bedarf weiterer Untersuchung.

Neben den individuellen Unterschieden zwischen Probanden hängt das Auftreten von Verhaltensänderungen stark von der vorliegenden Aufgabensituation ab. Im beschriebenen Experiment lag immer die gleiche Zweiaufgabensituation vor, wodurch die Identifikation von beanspruchungsinduzierten Veränderungen stark vereinfacht ist. In reellen Einsätzen, vor allem im Bereich der militärischen Luftfahrzeugführung, variiert die vorliegende Aufgabensituation oft durch Veränderungen in der Umgebung des Mensch-Maschine Gesamtsystems. Beispielsweise kann das Auftreten der Strategie *Umriorisieren* durch einen Warndialog mit hoher Salienz ausgelöst werden, ohne dabei eine erhöhte Beanspruchungssituation zu signalisieren.

Im Folgenden wird das Nutzen der verhaltensbasierten Messungen für adaptive Automation Systeme anhand der in Kapitel 0 vorgestellten Bewertungskriterien genauer betrachtet.

Robustheit

Obwohl das Auftreten von Verhaltensänderungen einen Indikator für erhöhte Beanspruchung darstellt, treten diese auch gelegentlich in Situationen niedrigerer Beanspruchung auf. Aus diesem Grund reicht das Erkennen von Verhaltensänderungen nicht aus, um eine erhöhte Beanspruchungssituation zu identifizieren. Die beschriebene Möglichkeit für Fehlklassifikationen aufgrund von Situationsänderungen verringert zusätzlich die Aussagekraft dieses Beanspruchungsmaßes. Daher ist die Wertung der Robustheit von Verhaltensbasierten Messungen **mittelmäßig** bei einer vordefinierten Aufgabensituation und **niedrig** in einem veränderlichen Mehraufgabenkontext.

Sensitivität

Der Einsatz von Verhaltensbasierten Messungen für adaptive Automation bietet die Möglichkeit einer dreistufigen Unterscheidung von Beanspruchung. Hierbei wird eine niedrige Beanspruchung durch kein oder nur seltenes Auftreten von vereinfachenden Strategien. Bei häufigem Auftreten der Strategien *Lockerung der Höhenvorgabe* und *Verzögerung der Klassifizierungsaufgabe* kann von einer mittleren Beanspruchung ausgegangen werden und bei dem Einsatz der anderen Strategien zeichnet sich eine Überbeanspruchung ab. Gegenüber den kontextfreien Messmethoden stellt die ermittelte Beanspruchung hierbei einen direkten Indikator für notwendige Unterstützung dar anstatt eines zusätzlichen Messwerts bei dem ein Schwellwert festgelegt werden muss. Dadurch ist die Sensitivität insgesamt **hoch**, sofern die Aufgabensituation das Erkennen von Verhaltensänderungen zulässt.

Diagnostizität

Die Anwendbarkeit der Methode ist stark von der gegebenen Aufgabensituation abhängig. Aufgrund dieses Zusammenhangs kann auf Basis der zugrundeliegenden Aufgaben bestimmt werden welche Art von Beanspruchung der gegebenen Situation zugrunde liegt,

indem die Aufgaben beispielsweise mit Hilfe der VACP Dimensionen analysiert wird. Über diese Information hinaus bietet die Ermittlungsmethode keine zusätzliche Unterscheidung über die Art der Beanspruchung und die Diagnostizität ist **sehr niedrig**.

Individualität

Im vorliegenden Experiment haben die Probanden mit weniger Selbstadaptiven Strategien grundsätzlich eine bessere Leistung erbracht. Aufgrund dessen liegt die Annahme nahe, dass die anderen Probanden nicht ausreichend beansprucht waren, um den Einsatz von selbstadaptiven Strategien zu benötigen. Dementsprechend signalisiert das Auftreten von entsprechenden Verhaltensänderungen die individuelle Notwendigkeit für Unterstützungsfunktionen, wodurch die Individualität als **hoch** einzustufen ist.

Generalisierbarkeit

Die beobachteten Strategien können potenziell verallgemeinert werden zu dem *Vereinfachen einer Aufgabe*, dem *Verzögern der Aufgabendurchführung*, dem *Multitasking* und dem *gezielten Kontextwechseln*. Diese Allgemeineren Strategien lassen sich daher kontextunabhängig definieren. Dennoch ist die Möglichkeit Strategien anzuwenden von den jeweiligen Aufgabensituationen abhängig, weshalb für den Einsatz der Messmethode eine domänenspezifische Modellbildung der erwarteten Aufgabensituationen notwendig ist. Zusätzlich muss dem Messsystem ein Modell zur Identifikation der jeweiligen Situationen mitgeteilt werden, damit es die Anwendungen der jeweiligen Strategien ermitteln kann. Daher ist die Generalisierbarkeit der Messmethode **sehr niedrig**.

Direkte Anwendbarkeit

Die beschriebenen Strategien teilen neben der Beanspruchung auch die Aufgabe mit in der sie auftreten. Beispielsweise wurde die *Verzögerung der Aufklärungsaufgabe* vorwiegend eingesetzt, wenn die Schwierigkeit der Flugaufgabe angestiegen ist. Daher können die unterschiedlichen Strategien einen Indikator dafür darstellen, welche Aufgabe die Ursache einer hohen Beanspruchung darstellt. Da die Erkennung der Verhaltensänderungen allerdings immer noch auch eine Analyse der Aufgabensituation voraussetzt ist die direkte Anwendbarkeit insgesamt nur **mittelmäßig**.

4.4.2 Modellbasiertes Messverfahren

Bei dem Experiment zum modellbasierten Messverfahren sollte ermittelt werden, inwiefern Domänenwissen genutzt werden kann, um den mentalen Zustand eines Piloten genauer zu erfassen und somit die Eingriffsmöglichkeiten für adaptive Automation genauer zu spezifizieren. Hierbei soll insbesondere die Möglichkeit untersucht werden, ob der Aufgabenkontext des Probanden ermittelt werden kann, um die Reaktion des adaptiven

Systems auf tatsächlich durchgeführte Aufgaben zu ermöglichen, anstatt auf Erwartungen zu beruhen welche Aufgaben durchgeführt werden müssten.

4.4.2.1 Experimentaldesign

Da es in der Untersuchung um die Erkennung und Nutzbarmachung von realistischen Aufgabensituationen gehen soll, wurde in der Gestaltung des Experiments ein besonderes Augenmerk auf die Abbildung von ebenso realistischen Aufgabensituationen gelegt. Die entstandenen Experimentalmissionen basieren auf einer Analyse des Future Operating Environments des ‚European Technology Acquisition Programme (ETAP)‘, also einer Vorhersage für zukünftige Militärmissionen, die auch als Grundlage für das nächste Kampfflugzeug dient. Nach der Festlegung auf zu erprobende Missionstypen wurden die Experimente mit Experten der deutschen Luftwaffe diskutiert, damit der notwendige Grad an Realismus erreicht werden konnte.

4.4.2.2 Beschreibung

Innerhalb des Experiments wurden sechs militärische Kampfflugzeug Missionen geflogen, in denen der Proband der Führer eines MUM-T Verbunds ist. Die Missionen wurden in drei Missionsgruppen unterteilt, zwischen denen die erwartete Stärke der Feindkräfte variiert wurde, während die zwei Missionen innerhalb einer Missionsgruppe jeweils einen ähnlichen Missionsablauf abbildeten. Jede Mission wurde mit einem Briefing eingeleitet, indem die Details der jeweiligen Mission vorgestellt wurden und ein Plan vorgegeben wurde, mit dem die Missionsziele erreicht werden können. Dieser initiale Plan wurde den Probanden bereits in die Simulationsumgebung eingetragen. Allerdings wurde den Probanden trotzdem die Freiheit eingeräumt den vorgegebenen Plan jederzeit anzupassen, damit sie ihr Domänenwissen gezielt einsetzen können.

MV-A

Die ersten beiden Missionen basieren auf einem asymmetrischen Konflikt, bei dem die eigenen Kräfte grundsätzlich die Luftüberlegenheit haben und somit kein Kontakt mit feindlichen Luftfahrzeugen zu erwarten ist. Die simulierte Konfliktlage besteht dabei darin, dass die geflogene Mission sich gegen eine Terrorgruppierung (Kiam) in einem fiktiven Land (Lorea) richtet. Aus diesem Grund liegt ein Fokus für die Probanden darauf die Mission unentdeckt durchzuführen, um keinen Konflikt mit den Loreanern zu erzeugen und den Kontakt mit feindlichen Flugzeugen zu vermeiden. Ziel der Missionen war es in einem vordefinierten Gebiet ein Terroristencamp zu suchen, auf Basis von elektrooptischem Bildmaterial zu verifizieren, dass es sich um das richtige Camp handelt und es abschließend zu bekämpfen. Als zusätzliches Ziel wurden die Probanden dazu angehalten ein weiteres Foto des Camps nach der Bekämpfung aufzunehmen, um den Erfolg der Bekämpfung zu

bewerten, sofern dadurch kein Kontakt mit anderen Luftfahrzeugen entsteht. Als zusätzliche Rahmenbedingung wurde den Probanden mitgeteilt, dass das Eindringen in das Zielgebiet die Loreaner alarmieren wird und im Rahmen eines *Quick Response Alert* (QRA) Flugzeuge zur Untersuchung losgeschickt werden, die fünf Minuten nach dem Auslösen eintreffen.

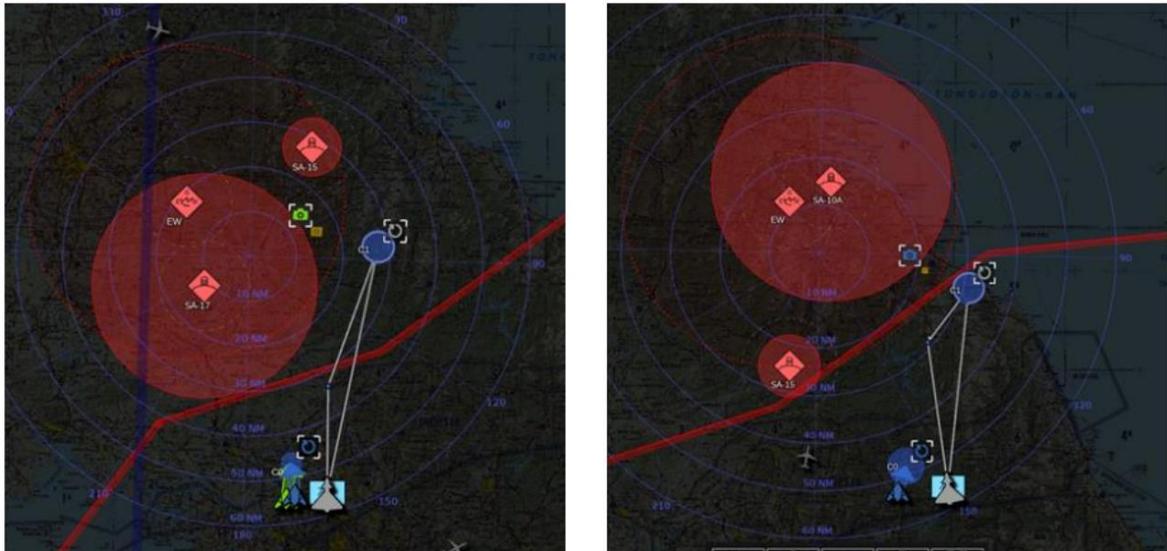


Abbildung 50: MV-A im modellbasierten Experiment

Abbildung 50 stellt den vorgefertigten Plan für diese beiden Missionen in unserer taktischen Kartenanzeige dar. In beiden Missionen bestehen die bekannten Feindkräfte aus einer bodengebundenen Flugabwehrstation (SAM-Site) mit einer geringen Reichweite, einer SAM-Site mit hoher Reichweite und einer zusätzlichen Radarstellung. Aufgrund der Aufklärungslage wird davon ausgegangen, dass keine weiteren unbekanntes Stellungen existieren. Der Plan besteht in beiden Missionsvarianten daraus, dass das bemannte Flugzeug in eine ausgewiesenes Halteareal fliegt, um dort in der Nähe des Einsatzgebietes abrufbar zu sein, während ein unbemanntes Flugzeug das Ziel innerhalb des Einsatzgebietes sucht. Dabei ist das Einsatzgebiet jeweils im Bereich der Radarstellung, aber außerhalb der SAM-Sites. Somit starten die fünf Minuten bis zur Ankunft der QRA-Flugzeuge beim Einfliegen eines UAV in die Radarstellung. Innerhalb dieses Zeitraums sucht das UAV im Zielgebiet nach dem Terroristencamp und generiert Fotos von potenziellen Zielen, die vom Probanden während des Flugs klassifiziert werden. Nach einer positiven Identifikation wird daraufhin die Bekämpfung des Ziels geplant und geflogen, bevor der Rückflug angetreten wird. Je nach Dauer dieses Prozesses kann der Proband auf dem Rückflug noch veranlassen, dass eines der UAV ein Foto zur Verifizierung der Bekämpfungsleistung erzeugt, bevor die QRA-Flugzeuge eintreffen.

MV-B

Das zweite Paar von Missionen beschreibt einen Konflikt mit einem stärkeren Gegner, bei dem die Bekämpfung von SAM-Sites sowie feindlichen Flugzeugen notwendig ist. Neben wenigen bekannten SAM-Sites wird bei diesen Missionen auch von sogenannten Pop-up SAM-Sites ausgegangen. Hierbei handelt es sich um SAM-Sites, die zuvor nicht bekannt sind und somit zu jedem Zeitpunkt während der Mission als zusätzliche Bedrohung auftreten können. Ziel der Missionen ist erneut die Verifikation und Bekämpfung einer bodengebundenen Raketenabschussvorrichtung in einem vorgegebenen Zielgebiet.

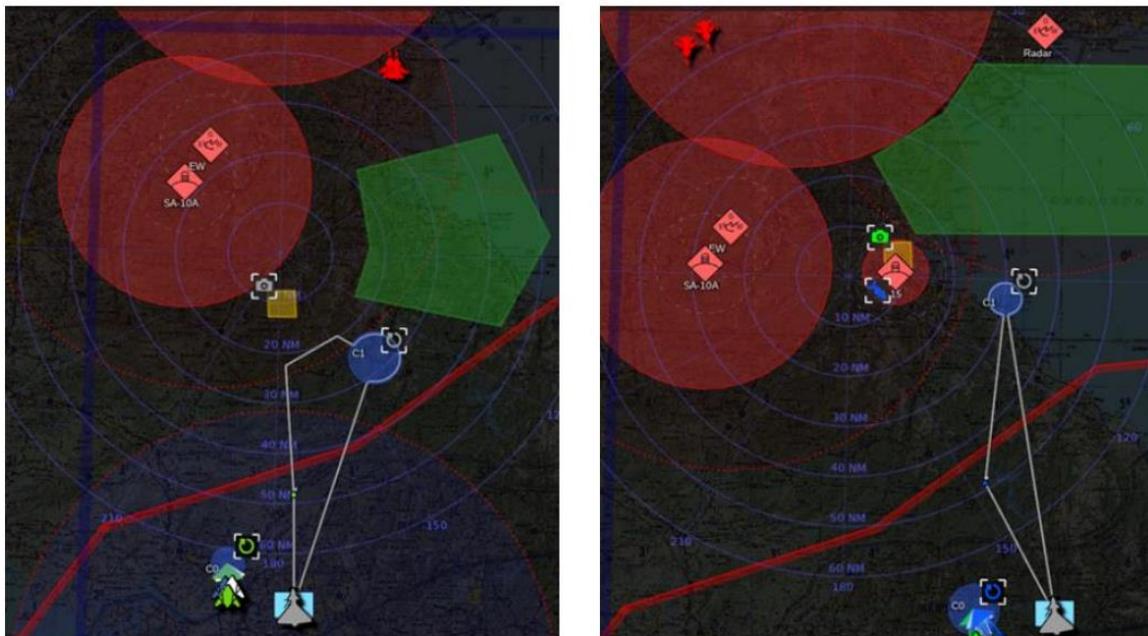


Abbildung 51: MV-B im modellbasierten Experiment

Die beiden Missionsgebiete werden in Abbildung 51 dargestellt. Die Bedrohungslage stellt sich dabei ähnlich zu der Lage aus den MV-A Missionen dar. Allerdings erzeugen die feindlichen Flugzeuge, sowie die Erwartung von Pop-up SAM-Sites ein stark erhöhtes Bedrohungspotential. Die zusätzlichen grünen Polygone (Air-to-Air Area) in der Kartendarstellung wurden im Briefing vordefiniert, damit sie einem UAV für die Bekämpfung von Feindflugzeugen zugewiesen werden können. Der vorgefertigte Plan besteht in diesen beiden Missionen erneut darin, dass das bemannte Flugzeug sich in ein Halteareal begibt und von dort aus vorwiegend die unbemannten Kräfte befehligt. Währenddessen wird ein UAV damit beauftragt das Ziel innerhalb des Zielgebiets ausfindig zu machen und ein Foto an den Probanden zu übermitteln. Dazu ist eine Bekämpfung einer SAM-Site notwendig, in deren Wirkungsbereich das Zielgebiet liegt. Diese Bekämpfung konnte allerdings nur in Mission MV-B-2 vorgeplant werden, da die SAM-Site in MV-B-1 dem Probanden nicht bekannt war und somit eine dynamische Umplanung während der Mission vorgenommen werden musste. Einem weiteren UAV musste mithilfe der Air-to-Air

Area die Bekämpfung von Feindflugzeugen zugewiesen werden. Hierbei ist es ebenfalls möglich die Air-to-Air Area zu verschieben oder durch eine neue zu ersetzen. Die beiden Bekämpfungsaufgaben sorgen in der Mission für viel Dynamik, weil die Bekämpfungen fehlschlagen können und somit der Plan während der Missionsdurchführung angepasst werden muss. Auch in diesen Missionen wird nach der Verifizierung der Ziele eine Bekämpfung durchgeführt und auf dem Rückweg ein weiteres Foto zur Abschätzung der Bekämpfungsleistung erzeugt.

MV-C

Die letzten zwei Missionen bilden jeweils eine ‚Deep strike‘ Mission ab, bei der weit in das Hoheitsgebiet einer feindlichen Luftabwehr vorgedrungen werden muss, um dort ein Ziel zu bekämpfen. Die feindlichen Kräfte bestehen hierbei aus einer Vielzahl von bekannten und unbekannt SAM-Sites, sowie mehreren feindlichen Flugzeugen. Aufgrund der Bedrohungslage war es nötig nah über dem Grund zu fliegen, damit das vorliegende Gelände die Flugzeuge vor feindlichen SAM-Sites abschirmen kann. Dies führt zu einer erhöhten Belastung, da der implementierte Autopilot nicht in der Lage ist den entsprechenden Geländefolgeflug durchzuführen und die Probanden somit vorwiegend manuell fliegen mussten.

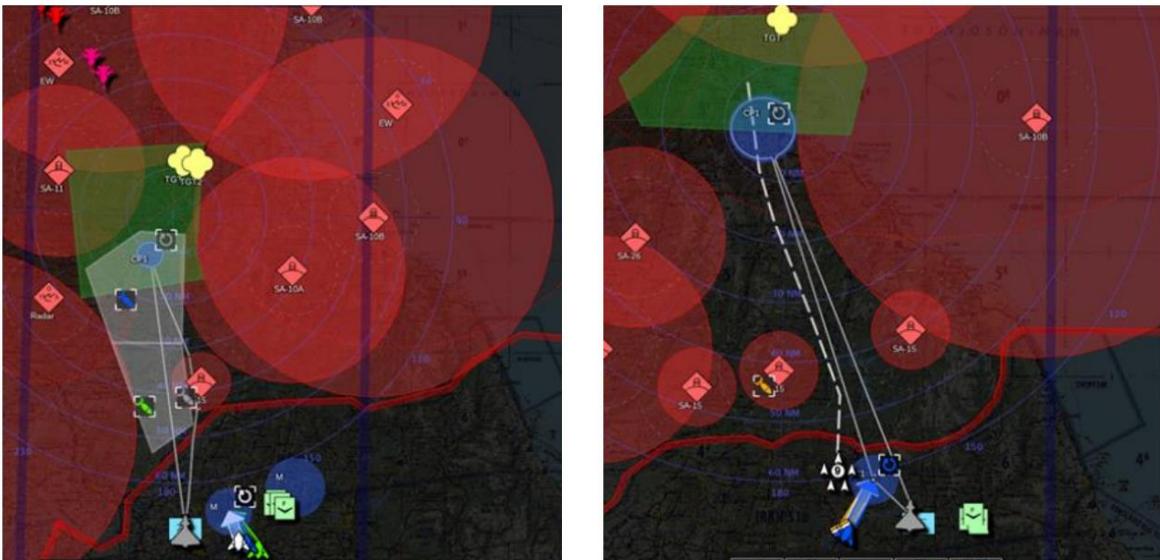


Abbildung 52: MV-C im modellbasierten Experiment

Die zwei, in Abbildung 52 dargestellten, Varianten der Mission MV-C hatten als zusätzliches Rahmenkriterium eine weitere Gruppe von Flugzeugen (hellgrüne Rechtecke in den Teilabbildungen unten). Diese Flugzeuge sind verbündete Kräfte; die im Zielgebiet einen Bunker bekämpfen sollen, weshalb sie von dem Probanden und seinen unbemannten Flugzeugen ins Zielgebiet eskortiert werden müssen. Dazu musste der Proband seine Kräfte in einem Umkreis von 35 nautischen Meilen zu den verbündeten Flugzeugen halten, um neu

auftretende Bedrohungen schnellstmöglich bekämpfen zu können. Trotz dessen müssen die befehligten Kräfte voraus fliegen, um bereits bekannte Bedrohungen vorab zu bekämpfen und unbekannte Bedrohungen aufzudecken.

Die zwei Varianten der Mission unterscheiden sich hierbei durch die zur Verfügung stehenden Kräfte. In der Variante MV-C-2 (rechts in Abbildung 52) steht zusätzlich zu den unbemannten Flugzeugen noch ein Schwarm zur Verfügung, der voraus fliegt um SAM-Sites frühzeitig aufzudecken, damit diese explizit zur Bekämpfung zugewiesen werden können. Demgegenüber wurden in der Variante MV-C-1 (links in der Abbildung) den unbemannten mehr Freiheiten eingeräumt. Dazu wurden analog zu den grünen Air-to-Air Areas weiße SEAD (Suppression of enemy air defenses) Areas (weiße Polygone in der Abbildung) als taktische Elemente hinzugefügt. Sobald einem UAV eine solche Area zugewiesen wurde, musste es auftretende SAM-Sites in diesem Bereich selbstständig unterdrücken und konnte dadurch SAM-Sites bekämpfen, ohne dass sie explizit zugewiesen wurden.

4.4.2.3 Durchführung

Die Experimente wurden mit denselben acht Kampfflugzeugpiloten der Luftwaffe durchgeführt, wie bereits im augenbasierten Experiment (siehe Tabelle 4). Jeder der Piloten durchlief dabei ein zweitägiges Training in dem MUM-T Missionssimulator, bevor die Experimentalmissionen geflogen wurden. Zu Beginn eines Experimentaltags wurden den Probanden die BioPac Bionomadix Sensoren angelegt, um die EKG-Daten der Probanden über den Tag zu ermitteln und eine 10-minütige Baseline-Messung durchgeführt. Zusätzlich wurden die Kameras des Blickbewegungsmesssystems auf die Probanden kalibriert.

Daraufhin erhielten die Probanden eine Beschreibung des militärischen Rahmenkonflikts in dem die Experimentalmissionen geflogen wurden. Während des Experiments wurde dieser Rahmenkonflikt genutzt, um die zunehmende Stärke der feindlichen Kräfte zu Begründen. Vor jeder Experimentalmission erhielten die Probanden eine Einführung in die jeweilige Mission. Hierbei wurde auf das Missionsgebiet, die zur Verfügung stehenden Kräfte, sowie deren Bewaffnung, das Missionsziel und den vorgefertigten Plan eingegangen.

Während der Experimentalmissionen wurden die Herzrattendaten, sowie die visuellen und manuellen Interaktionen der Probanden mit der Mensch-Maschine Schnittstelle aufgezeichnet. Mit Hilfe dieser Interaktionen wurde die Prozesskette der Aktivitätsbestimmung aus Kapitel 4.2.5 genutzt, um zu jedem Zeitpunkt eine Schätzung über die Aktivität des Piloten zu generieren. Zusätzlich wurde ein Abbild der gesamten Simulation mit einer zeitlichen Auflösung von 20 Millisekunden gespeichert.

Nach jedem Experiment wurde die Aufzeichnung der Mission abgespielt und der Proband wurde gebeten den Missionsablauf und die dabei getroffenen Entscheidungen zu

kommentieren. Hierbei wurde insbesondere darauf eingegangen, wodurch die Beanspruchung in den schwierigen Phasen ausgelöst wurde.

4.4.2.4 Ergebnisse

Mithilfe der Daten aus der zuvor beschriebenen Aktivitätserkennung wurden Aktivitätsprofile für die einzelnen Probanden über die Missionsverläufe erzeugt. Dazu wurden die modellierten Aufgaben, wie in (Lindner et al. 2022) beschrieben, in unterschiedliche Kategorien eingeteilt. Dies war notwendig da die einzelnen erkannten Aufgaben auf Elemente in der Situationsumgebung bezogen sind. Insbesondere in den Missionen des Missionstyps MV-C treten viele SAM-Sites und Feindflugzeuge auf, wodurch Aufgaben wie ‚Threat Assessment‘, bei der ein Pilot die Bedrohung durch eine SAM-Site abschätzt, in einem Zeitschritt mehrfach instanziiert werden und die Bedeutung der einzelnen Aufgaben abnimmt.

Die Aktivitätsprofile beschreiben für jede Minute den Prozentanteil der Aufgabenkategorien, die ein Proband in diesem Zeitraum durchgeführt hat. Hierbei ist es möglich, dass die Summe der Prozentanteile 100% übersteigt, da die Aktivitätserkennung mehrere unterschiedliche Aufgaben gleichzeitig als aktiv ansehen kann. Abbildung 53 stellt zwei Aktivitätsprofile für die zwei Missionen des Missionstyps MV-A dar.

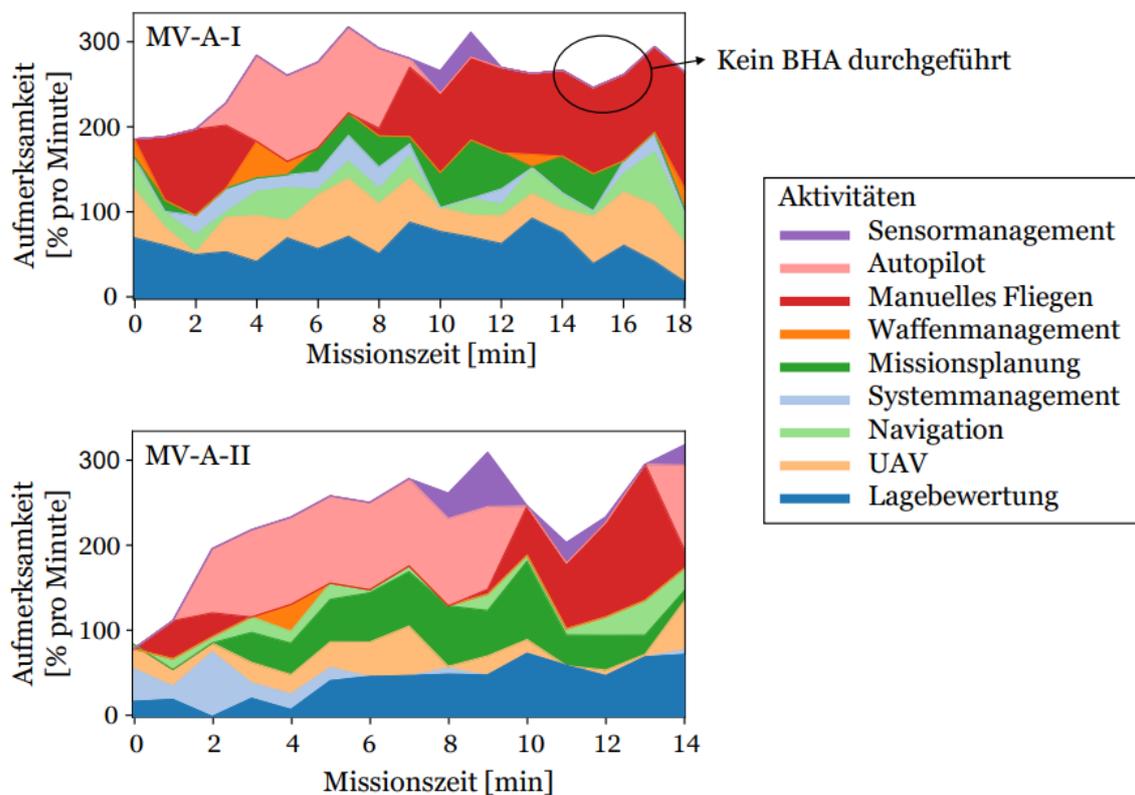


Abbildung 53: Durchschnittliche Aufgabenverteilung der Piloten über die Missionszeit

Die Aktivitätsprofile zeigen den Mehrwert einer Aktivitätserkennung gegenüber situationsadaptiver Automation. Im oberen Aktivitätsprofil ist zu sehen, dass der Proband das ‚*Battle Hit Assessment*‘ (BHA), also die Verifikation der Zerstörung des Ziels, nicht durchgeführt hat. In einer situationsadaptiven Automation würde hierbei die Frage auftreten, ob die Aufgabe nicht durchgeführt wurde oder lediglich das Ergebnis der Aufgabendurchführung dem System nicht mitgeteilt wurde. Ein ähnlicher Fall befindet sich in dem unteren Aktivitätsprofil, bei dem der Pilot vor der Bekämpfung (Minute 10) sich nicht mit der Konfiguration seiner Waffen beschäftigt. Sind die Waffen hierbei nicht aktiviert stellt dies einen Fall dar, bei dem ein Hinweis notwendig ist, der dem Piloten mitteilt, dass der er seinen Angriff nicht fliegen kann. Im Gegensatz dazu kann im oberen Fall die Meldung unterdrückt oder angepasst werden, weil der Pilot die ‚Waffenmanagement‘ Aufgabe durchführt

Die Bestimmung des Aufgabenkontexts stellt somit eine wertvolle Erweiterung neben der Beanspruchung dar. Sie kann genutzt werden, um in kritischen Situationen die Ursache der Überlastung genauer zu verstehen oder überflüssige Warnungen zu vermeiden.

Die Nachbesprechung mit den Probanden hat ergeben, dass die Hauptursachen für eine hohe Beanspruchung unerwartetes Verhalten der UAV oder unvorhergesehene Änderungen in der Missionsdynamik waren. Das unerwartete Verhalten der UAV war hierbei häufig auf fehlerhafte Eingaben in der Aufgabenzuweisung zurückzuführen. Ursache hierfür war häufig, dass die Probanden dachten sie hätten einem UAV eine Aufgabe zugewiesen, dabei aber vergessen haben die Aufgabenzuweisung zu bestätigen. Dadurch hat sich gezeigt, dass die modellbasierte Beanspruchungsmessung nur bedingt eignet, um aktive Überlastungen zu erkennen, sondern eignet sich maximal zur Vorhersage zukünftiger kritischer Situationen.

Im Gegensatz dazu kann die Herzrate genauer analysiert werden, wenn sie mit Domänenwissen gemeinsam betrachtet wird. Abbildung 54 zeigt zwei Herzratenverläufe in den Missionen des Typs MV-A oben und einen Verlauf der Mission MV-C-VI unten. Die Herzdaten in MV-A sind dabei sehr konstant im Vergleich zu dem Verlauf in MV-C-VI. Dies ist insbesondere auch der Fall, für das Auftreten der QRA aufgrund des Verstreichens der fünf Minuten nach dem Eintritt in die feindlichen Sensoren. Innerhalb der Befragung teilte der Pilot mit, dass ihm die Feindflugzeuge zwar bewusst waren, sie aber gleichzeitig als irrelevant bewertet wurden, weil die Distanz zwischen den eigenen Flugzeugen und der QRA ausreichend groß war. Der untere Verlauf wiederum zeigt einen starken Anstieg innerhalb der violett gekennzeichneten Bereiche, die den Zeitbereich nach dem Auftauchen einer neuen *SAM-Site* anzeigen. Von diesen Pop-up Bedrohungen war der Pilot nach eigenen Angaben direkt betroffen und kam dadurch in eine beanspruchende Situation, in der er sich Unterstützung durch den automatisierten Planer gewünscht hätte, indem dieser

beispielsweise einen neuen Plan generiert indem ein UAV direkt damit beauftragt wird die SAM-Site zu bekämpfen

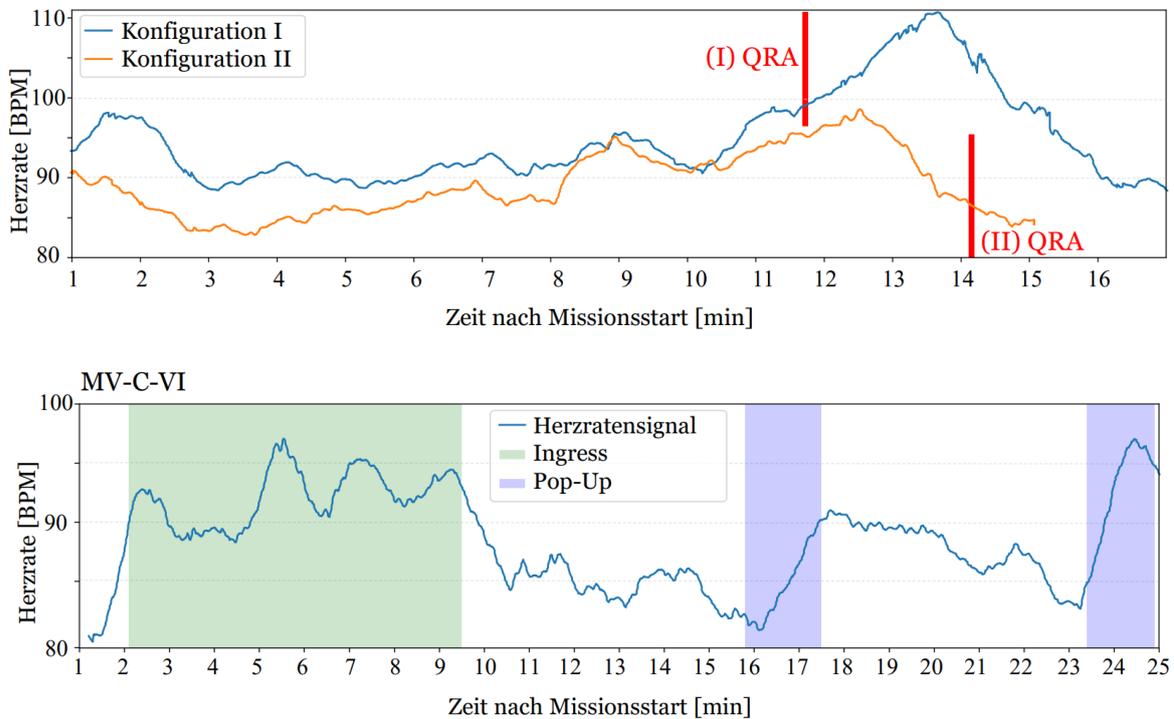


Abbildung 54: Vergleich der Herzratenänderung zwischen den leichten Missionen (oben) und einer schweren Mission (unten)

4.4.2.5 Bewertung

Der modellbasierte Ansatz liefert eine Schätzung der durchgeführten Aufgaben, wodurch eine adaptive Automation passend ausgewählt werden kann. Die Schwachstellen liegen allerdings in der Beschreibung der Beanspruchung. Zum einen kann die modellierte Beanspruchung nicht auf einen individuellen Nutzer ausgelegt werden, sondern spiegelt einen generischen Nutzer wider. Zum anderen hat die Diskussion innerhalb des Experiments gezeigt, dass Beanspruchungssituationen nur schwer durch die Aktivität beschreibbar ist und stattdessen durch unterschiedliche Facetten bei der gleichen Aktivität entstehen. Eine theoretische Möglichkeit besteht hier darin, dass die Facetten in den einzelnen Aufgaben voneinander abgegrenzt werden und durch eine genauere Modellierung erneut jede Möglichkeit abgedeckt werden kann.

Robustheit

Die Erkennung der Beanspruchung auf Basis einer Beanspruchungsmodellierung kann auf Basis der Interaktionen durchgeführt werden. Bereits die manuellen Interaktionen liefern ein Abbild der durchgeführten Aufgaben und können dadurch für eine Schätzung der

Beanspruchung genutzt werden. Die Daten der Blickbewegungsmessung können weiterhin genutzt werden, um das Beanspruchungsprofil noch weiter zu präzisieren. Insgesamt ist die Robustheit der Methode **sehr hoch**.

Sensitivität

Durch die Möglichkeit zur Verwendung beliebiger Beanspruchungsmodelle kann die ermittelte Beanspruchung an die jeweilige Domäne angepasst werden. Beispielsweise kann berücksichtigt werden, wenn eine Mehraufgabensituation nicht möglich ist, indem ein Beanspruchungsmodell speziell für diesen Anwendungsfall gewählt wird. Zusätzlich kann eine verbesserte Unterscheidung einzelner Beanspruchungsstufen erzeugt werden, indem eine modellierte Aufgabe aufgeteilt wird, um zwei unterschiedlich beanspruchende Varianten der Aufgabe zu unterscheiden. Aufgrund der notwendigen Modellierung ist der Detailtiefe allerdings eine Grenze gesetzt, wodurch die Sensitivität der Messmethode nur **hoch** ist.

Diagnostizität

Im Vergleich zur Sensitivität ist die Diagnostizität einfacher zu modellieren. Für gewöhnlich beansprucht eine schwierigere Variante der gleichen Aufgabe dieselbe Ressource wie eine einfache Variante. Dadurch ist eine genaue Aufschlüsselung der einzelnen Aufgaben für die Diagnostizität weniger relevant als für die Sensitivität. Da ansonsten die gleiche Argumentation gilt, ist die Diagnostizität des modellbasierten Verfahrens **sehr hoch**.

Individualität

Da die Beanspruchungsmodellierung aus diesem Experiment unabhängig von den Piloten entstehen musste, konnte auf individuelle Eigenschaften der einzelnen Piloten nicht eingegangen werden, sondern die Beanspruchung der Aufgabe wurde auf einen allgemeinen Piloten ausgelegt. Wie in (Mund et al. 2020) diskutiert wurde kann dieser fehlenden Individualität mithilfe einer Kombination von maschinellen Lernverfahren und kontext-freien Beanspruchungsmessungen entgegengewirkt werden, allerdings benötigt ein solcher Ansatz viele Daten des jeweiligen Piloten. Insgesamt ist die Individualität daher **niedrig**.

Generalisierbarkeit

Die Modellierung der möglichen Aufgaben inklusive der dafür notwendigen Interaktionen und der damit verbundenen Beanspruchung setzt ein sehr genaues Verständnis der modellierten Domäne voraus. Diese Notwendigkeit sorgt für eine **sehr niedrige** Generalisierbarkeit der Messmethode.

Direkte Anwendbarkeit

Die beschriebene Methode liefert neben einer Beanspruchungsschätzung zusätzlich die dabei bearbeiteten Aufgaben. Dadurch wird neben der Entdeckung einer extremen

Beanspruchungssituation auch gleichzeitig die Ursache ermittelt. Eine darauf aufbauende adaptive Automation erhält somit einen möglichen Eingriffspunkt für eine automatisierte Unterstützung, wodurch die direkte Anwendbarkeit **sehr hoch** ist.

5. Diskussion

Jede der vorgestellten Messmethoden hat seine eigenen Vor- und Nachteile. Wie erwartet liegen die Vorteile der kontextfreien Methoden in der Individualität und der Generalisierbarkeit, während die Stärken der kontextbehafteten Methoden in der direkten Anwendbarkeit und dem Domänenwissen liegen. Bei einem genaueren Blick in die einzelnen Varianten bietet die EKG-basierte Messmethode den Vorteil schnell einsetzbar zu sein und wenig Kalibrieraufwand zu benötigen, während die kontextfreien Augenmessungen eine zusätzliche Ermittlung der Aufmerksamkeit mitliefert, dafür allerdings ausschließlich an statischen Arbeitsplätzen eingesetzt werden kann. Im Bereich der kontextbehafteten Messungen kann die verhaltensbasierte Messung bereits vor dem Einbruch einem Leistungseinbruch individuelle Situationen hoher Beanspruchung identifizieren, kann dafür allerdings nur eingesetzt werden, wenn gleichartige Aufgabenkontexte gehäuft auftreten. Die modellbasierte Methode kann vielseitig eingesetzt werden, da die Echtzeiterkennung im einfachsten Fall ohne zusätzliche Sensoren auskommt, indem sie sich ausschließlich auf die manuellen Interaktionen stützt. Allerdings benötigt die Modellierung ein sehr gutes Domänenwissen, wodurch der Einsatz der Methode aufwändig ist.

Für Aufgabenfelder mit einem hohen Risiko der getroffenen Entscheidungen empfiehlt sich eine Kombination einer kontextbehafteten und einer kontextfreien Beanspruchungsmessmethode. Im Rahmen dieser Arbeit wurde für die MUM-T Testapplikation auf Basis der gewonnenen Erkenntnisse ein adaptives Automationssystem konzipiert, welches die Vorteile der jeweiligen Messmethoden gewinnbringend einsetzen kann. Ziel war es dabei die Beanspruchung des Piloten in einem adäquaten Bereich zu halten indem konkurrierende Problemsituationen verglichen werden, damit der Pilot sich immer der wichtigsten Aufgabe widmen kann.

Die Lösung unterscheidet die Auslöser für automatisierte Unterstützungsfunktionen in die zwei Kategorien ‚Fehlzustand‘ und ‚Statusmeldung‘. Für beide Kategorien existiert jeweils ein Entscheidungsbaum, dessen Blätter die Auswahl einer Unterstützungsart darstellen. Die Entscheidungsknoten des Baums evaluieren zu jedem Auslöser den mentalen Zustand des Piloten anhand der Informationen der einzelnen psychophysiologischen Messmethoden.

Der Entscheidungsbaum für eine Statusmeldung ist in Abbildung 55 dargestellt und beinhaltet lediglich zwei Entscheidungen die die von (Dorneich et al. 2012) beschriebene Berücksichtigung von Etikette erzeugen soll. Unnötige Störungen während der Aufgabendurchführung können zu einer Leistungsminderung führen, weil der menschliche Operateur durch die Meldungen überlastet wird. Damit diese Überlastung vermieden werden kann, soll die Anzahl der unpassenden Statusmeldungen mithilfe des Entscheidungsbaums minimiert werden. Dazu wird durch das modellbasierte Messverfahren die Aktivität des Piloten geschätzt und evaluiert, ob die aktuelle Statusmeldung eine Verbindung zu den

durchgeführten Aufgaben hat. Ist dies der Fall, wird die Meldung an den Piloten weitergegeben, weil sie die Aufgabendurchführung unterstützen kann. Bearbeitet der Pilot allerdings keine verwandte Aufgabe wird überprüft wie hoch die Beanspruchung des Piloten ist. Für diese Überprüfung eignen sich die kontext-freien Beanspruchungsmessmethoden. Bei einer niedrigen Beanspruchung wird die Statusmeldung erneut an den Piloten weitergegeben, damit er den Überblick über die Situation behalten kann. Bei einer hohen Beanspruchung wird die Statusmeldung jedoch unterdrückt, wodurch der Pilot seine bestehende Aufgabe störungsfrei durchführen kann. Stattdessen wird die Meldung an eine Anzeige weitergeleitet, die der Pilot im nachhinein ansehen kann.

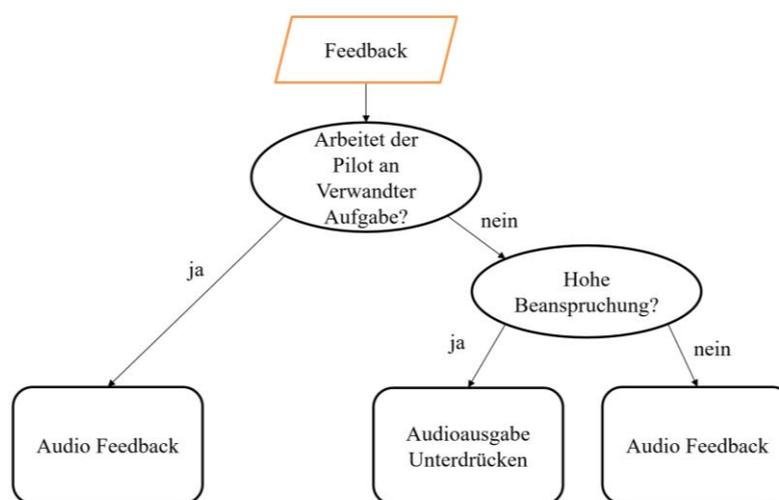


Abbildung 55: Entscheidungsbaum für den Umgang mit einer Statusmeldung

Im Falle der Erkennung eines Fehlzustands wird der Entscheidungsbaum in Abbildung 56 eingesetzt. Als Fehlzustand werden alle Zustände angesehen, in denen die Automation einen Handlungsbedarf identifiziert. Offensichtliche Fehlzustände sind daher unerwartete Bedrohungssituationen des Piloten, wie beispielsweise der unbewusste Flug in Richtung des Bodens. Allerdings kann auch eine erwartete Überlastungssituation, die ein automatisiertes System durch die Analyse des aktuellen Plans identifiziert hat, oder eine fehlende Zielverifikation durch den Piloten ein Fehlzustand sein.

Auch bei einem Fehlzustand wird zuerst überprüft, welcher Aktivität der Pilot nachgeht. Führt er eine Aufgabe durch, die mit der Behebung des Fehlzustands verbunden ist, wird davon ausgegangen, dass dem Piloten das Problem bewusst ist und er lediglich eine leichte Unterstützung benötigt. Daher wird mithilfe der Herzrate überprüft, ob die Beanspruchung in der nahen Vergangenheit stark gestiegen ist. Bei einem Beanspruchungsanstieg wird die Fehlerbehandlung durch mögliche Vereinfachungen der Aufgabe unterstützt. Im Falle einer Planungsaufgabe können beispielsweise Vorschläge von einem automatisierten Planer generiert werden. Ist stattdessen kein Beanspruchungsanstieg zu entdecken, wird

angenommen, dass der Pilot die Situation unter Kontrolle hat und somit keine Unterstützung notwendig ist.

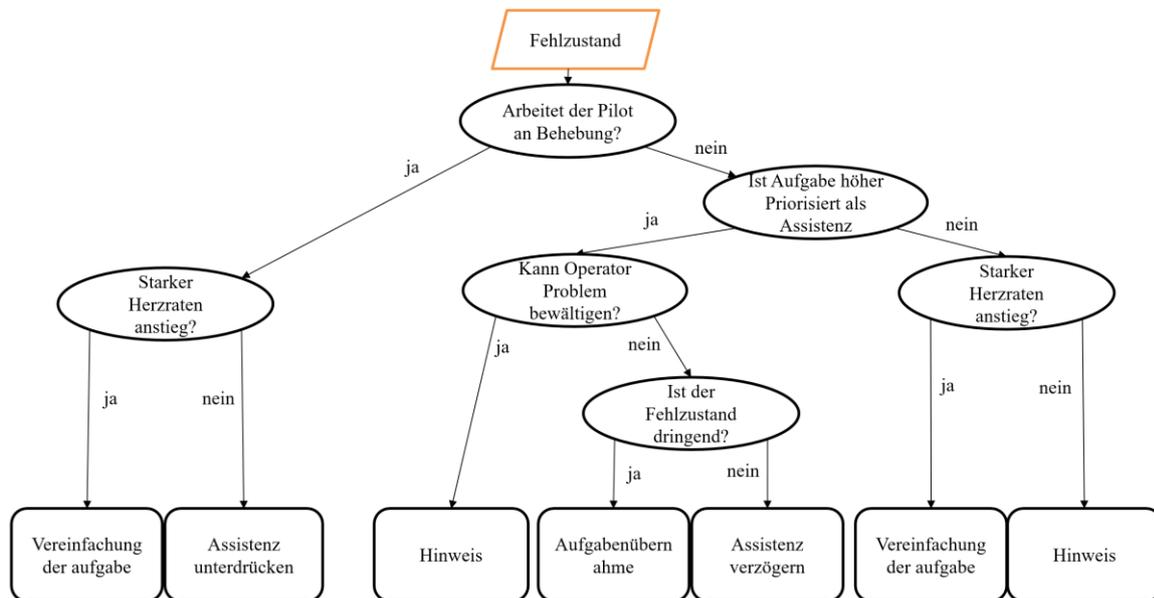


Abbildung 56: Entscheidungsbaum für den Umgang mit einem Fehlzustand

Führt der Pilot keine Aktion aus, die mit dem Fehlzustand in einem Zusammenhang steht, geht das adaptive System davon aus, dass dem Piloten der Fehlzustand nicht bewusst ist. In diesem Fall wird evaluiert, ob die derzeitige Aktivität eine höhere Priorität hat, damit unnötige Störmeldungen vermieden werden können. Bei einer höherpriorisierten Unterstützung wird erneut mithilfe der Herzrate die Beanspruchungsänderung bewertet. Ist kein solcher Anstieg zu erkennen wird der Pilot auf den Fehlzustand hingewiesen, da erwartet wird, dass der Pilot das Problem selbstständig beheben kann. Bei einer hohen Beanspruchung muss eine potenzielle Überlastung des Piloten in Kauf genommen werden. Um die Wahrscheinlichkeit der Überlastung minimal zu halten wird der Hinweis auf den Fehlzustand mit einem Lösungsvorschlag verknüpft.

Ist die Aktivität des Piloten höher priorisiert als der erkannte Fehlzustand muss die Wichtigkeit der Problemlösung evaluiert werden. Hierzu wird zuerst die modellbasierte Beanspruchungsschätzung eingesetzt, um die zu erwartende Gesamtbeanspruchung zu schätzen, die der Pilot empfindet, wenn er die Problembehandlung zusätzlich zu der aktuellen Aktivität ausführt. Übersteigt die kombinierte Beanspruchung die Belastungsgrenzen des Piloten nicht, kann ein Hinweis eingesetzt werden. Im Gegensatz dazu, wird bei einer Überlastung überprüft, ob die Behebung der Problemsituation verzögert werden kann. Ist dies nicht der Fall, kann bei besonders kritischen Fehlzuständen eine automatisierte Problembehebung in Erwägung gezogen werden. Für die meisten Anwendungsfälle wird allerdings auch hier eine vereinfachte Problembehandlung durch einen automatisierten Vorschlag angeboten.

6. Zusammenfassung und Ausblick

Diese Dissertation untersucht unterschiedliche Ansätze, die Automation innerhalb eines Mensch-Maschine-Systems an die Bedürfnisse des Bedieners anzupassen. Der Trend des verstärkten Einsatzes von Maschinen führt häufig dazu, dass der Mensch die übrigbleibenden Aufgabengebiete bearbeiten muss. Obwohl der Mensch eine hohe Flexibilität besitzt, führt die Notwendigkeit die Funktionslücken der Maschine zu kompensieren häufig dazu, dass der Bediener selbst keine optimale Leistung erbringen kann.

Diese automationsinduzierte Leistungssenkung führt häufig dazu, dass auch die Mensch-Maschine-Gesamtleistung abnimmt, obwohl die Maschine zusätzliche Funktionalitäten erhält. Beispielhafte Ursachen hierfür stellen Situationen dar in denen der Mensch sein Situationsbewusstsein verloren hat, weil er lange Zeit nicht mehr eingreifen musste und sich somit ablenken ließ. Außerdem führt eine höhere Leistungsfähigkeit der Maschine teilweise zu einem unangemessen hohen Vertrauen des Menschen in seine Automation, wodurch er dieser blind vertraut und dadurch erneut zusätzliche Fehler zulässt.

Durch menschadaptive Automation kann diesen Leistungssenkungen entgegengewirkt werden, indem die Maschine nicht immer alle Aufgaben tätigt, für die sie entwickelt wurde, sondern die Kontrolle dem Menschen überlässt, solange dieser dazu in der Lage ist. Auf Basis von Methoden der Beanspruchungsermittlung kann die Automation die Leistungsgrenzen des Menschen erkennen und möglichst nur dann unterstützend eingreifen, wenn es notwendig ist.

Im Rahmen dieser Arbeit wurden vier Experimente durchgeführt, die unterschiedliche Arten der Beanspruchungsermittlung einander gegenüberstellen sollten, damit die jeweiligen Vor- und Nachteile hervorgehoben werden können. Die Zielsetzung ist dabei ein adaptives Automationssystem für eine MUM-T Missionssimulationsumgebung zu schaffen, weshalb die Experimente in diesem Kontext stattgefunden haben.

Das entstandene Konzept erlaubt die Durchführung komplexer Kampfflugzeugmissionen, in denen ein Pilot mehrere unbemannte Kampfflugzeuge kommandiert, während er sich selbst im Feindgebiet befindet. Mithilfe eines maschinellen Verständnisses für die durchgeführten Aufgaben wird dabei versucht den Piloten bei der Ausführung seiner Aufgaben nur dann zu stören, wenn es zwingend erforderlich ist. Gleichzeitig werden durch den Einsatz kontextfreier Beanspruchungsmessmethoden individuelle Überlastungssituationen erkannt, damit die Automation stärker unterstützen kann.

Neben Evaluation des beschriebenen Konzepts und der Konzeptionierung eines solchen Systems für andere Domänen starker Mensch-Maschine-Kooperation kann die Ermittlung

von Unterstützungsbedarf auf einer Situationsbewusstseinsmodellierung ein vielversprechender Ansatz ein adaptives Automationssystem darstellen.

Der Einsatz dieser Systeme liefert die Möglichkeit die Leistungsfähigkeit von Mensch-Maschine-Systemen weiter auszureizen, indem die Anpassungsfähigkeit des Menschen an unerwartete Situationen optimal genutzt wird um neuartige Probleme frühzeitig zu beheben.

Abkürzungsverzeichnis

ACT-R	Adaptive Control of Thought—Rational
AS	Auditiv Räumlich (auditive spatial)
ATR	Automatisierte Zielerkennung (Automated Target Recognition)
AV	Auditiv Verbal
BHA	Battle Hit Assessment
BS	Blind Spot
COCOM	Contextual Control Model
CPC	Common Performance Conditions
CREAM	Cognitive Reliability and Error Analysis Method
CS	Kognitiv Räumlich (cognitive spatial)
CTL	Cognitive Task Load
CV	Kognitiv Verbal (cognitive verbal)
DASS	Defensive Aids Subsystem
DST	Dempster-Shafer Triplett
ECoG	Elektrokortikographie
EEG	Elektroenzephalographie
EKG	Elektrokardiograph
ETAP	European Technology Acquisition Programme
FAZ	Fast Approach Zone
FLOT	Forward line of own troops
fMRI	funktionelle Magnetresonanztomographie
fNIRS	funktionelle Nahinfrarot-Spektroskopie
FPZ	Front-Proximity Zone
HARM	High-speed Anti Radiation Missile
HF	High Frequency
HOTAS	Hands-On Throttle and Stick
HR	Herzrate
IFS	Institut für Flugsysteme
IPC	Interprozesskommunikation
JSON-RPC	JavaScript Object Notation - Remote Procedure Call
KVM	Keyboard, Video, Mouse
LOOCV	Leave-one-out cross validation
MABA-MABA	Men are better at Machines are better At
MAD	Median absolute difference
MEG	Magnetoenzephalographie
MHDD	Multifunctional Head-Down Display
MLP	Multilayer Perceptron
MUM-T	Manned-Unmanned Teaming
NAS	Network attached Storage

Abkürzungsverzeichnis

NATO	North Atlantic Treaty Organisation
PFD	Primary Flight Display
PKW	Personen Kraftwagen
QRA	Quick Response Alert
RASMUS	Real-Time Assessment of Multidimensional User State
RM	Reaktion Manuell
RMSSD	Root mean square of successive difference
RPZ	Rear-Proximity Zone
RR-Tri	Integral des RR-Interval Histograms geteilt durch seine Höhe
RV	Reaktion Verbal
SAGAT	Situational Awareness Global Assessment Technique
SAM	Surface to Air-Missile
SART	Situational Awareness Rating Technique
SD	Standardabweichung (Standard Deviation)
SEAD	Suppression of enemy air defenses
SHDD	Side Head-Down Display
SVM	Support Vector Machine
TLX	Task Load Index
UAV	Unbemanntes Flugzeug (Unmanned Aerial Vehicle)
VACP	Visuell, Auditiv, Kognitiv und Psychomotorisch
VS	Visuell Räumlich (visual spatial)
VV	Visuell Verbal

7. Literaturverzeichnis

- Ackerman, Phillip L. (1988): Determinants of individual differences during skill acquisition: Cognitive abilities and information processing. In: *Journal of Experimental Psychology: General* 117 (3), S. 288.
- Al-Rahayfeh, Amer; Faezipour, Miad (2013): Eye tracking and head movement detection: A state-of-art survey. In: *IEEE journal of translational engineering in health and medicine* 1, S. 2100212.
- Anderson, John R. (1991): Is human cognition adaptive?: na.
- Annett, John (2003): Hierarchical task analysis. In: *Handbook of cognitive task design* 2, S. 17–35.
- Bainbridge, Lisanne (1983): Ironies of automation. In: *Analysis, design and evaluation of man-machine systems*: Elsevier, S. 129–135.
- Beatty, Jackson (1982): Task-evoked pupillary responses, processing load, and the structure of processing resources. In: *Psychological bulletin* 91 (2), S. 276.
- Beatty, Jackson; Lucero-Wagoner, Brennis (2000): The pupillary system. In: *Handbook of psychophysiology* 2 (142-162).
- Beevis, D.; Bost, R.; Döring, B.; Nordø, E.; Oberman, F.; Papin, J. P. et al. (1994): Analysis techniques for man-machine system design. In: *NATO document AC/243 (Panel 8) TR/7 1* (2).
- Benedetto, Simone; Pedrotti, Marco; Minin, Luca; Baccino, Thierry; Re, Alessandra; Montanari, Roberto (2011): Driver workload and eye blink duration. In: *Transportation research part F: traffic psychology and behaviour* 14 (3), S. 199–208.
- Berthold, André; Jameson, Anthony (1999): Interpreting symptoms of cognitive load in speech input. In: *UM99 user modeling*: Springer, S. 235–244.
- Biau, Gérard; Scornet, Erwan (2016): A random forest guided tour. In: *Test* 25 (2), S. 197–227.
- Bierbaum, Carl R.; Szabo, Sandra M.; Aldrich, Theodore B.: Task Analysis of the UH-60 Mission and Decision Rules for Developing a UH-60 Workload Prediction Model. Volume 1. Summary Report. ANACAPA SCIENCES INC FORT RUCKER AL.
- BIOPAC Systems Inc. Online verfügbar unter <https://www.biopac.com/product/bionomadix-2ch-wireless-ecg-transmitter/>.

Boehne, Donna M.; Paese, Paul W. (2000): Deciding whether to complete or terminate an unfinished project: A strong test of the project completion hypothesis. In: *Organizational Behavior and Human Decision Processes* 81 (2), S. 178–194.

Bommer, Sharon Claxton; Fendley, Mary (2018): A theoretical framework for evaluating mental workload resources in human systems design for manufacturing operations. In: *International Journal of Industrial Ergonomics* 63, S. 7–17.

Borghini, Gianluca; Vecchiato, Giovanni; Toppi, Jlenia; Astolfi, Laura; Maglione, A.; Isabella, R. et al. (2012): Assessment of mental fatigue during car driving by using high resolution EEG activity and neurophysiologic indices. In: 2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society: IEEE, S. 6442–6445.

Bradski, Gary; Kaehler, Adrian (2000): OpenCV. In: *Dr. Dobb's journal of software tools* 3.

Brand, Yannick; Schulte, Axel (2017): Model-based prediction of workload for adaptive associate systems. In: 2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC): IEEE, S. 1722–1727.

Brand, Yannick; Schulte, Axel (2021): Workload-adaptive and task-specific support for cockpit crews: design and evaluation of an adaptive associate system. In: *Human-Intelligent Systems Integration* 3 (2), S. 187–199.

Broadbent, Donald Eric (1957): Immediate memory and simultaneous stimuli. In: *Quarterly Journal of Experimental Psychology* 9 (1), S. 1–11.

Brouwer, Anne-Marie; Hogervorst, Maarten A.; van Erp, Jan B. F.; Heffelaar, Tobias; Zimmerman, Patrick H.; Oostenveld, Robert (2012): Estimating workload using EEG spectral power and ERPs in the n-back task. In: *Journal of neural engineering* 9 (4), S. 45008.

Bustamante, Ernesto A.; Spain, Randall D. (2008): Measurement invariance of the Nasa TLX. In: Proceedings of the Human Factors and Ergonomics Society Annual Meeting: SAGE Publications Sage CA: Los Angeles, CA (52), S. 1522–1526.

Bye, Andreas; Hollnagel, Erik; Brendeford, Tor Steinar (1999): Human-machine function allocation: a functional modelling approach. In: *Reliability Engineering & System Safety* 64 (2), S. 291–300.

Calhoun, Gloria L.; Ruff, Heath A.; Spriggs, Sarah; Murray, Crystal (2012): Tailored performance-based adaptive levels of automation. In: Proceedings of the Human Factors and Ergonomics Society Annual Meeting: SAGE Publications Sage CA: Los Angeles, CA (56), S. 413–417.

- Calhoun, Gloria L.; Ward, Victoria B. R.; Ruff, Heath A. (2011): Performance-based adaptive automation for supervisory control. In: Proceedings of the Human Factors and Ergonomics Society Annual Meeting: SAGE Publications Sage CA: Los Angeles, CA (55), S. 2059–2063.
- Canham, Lyn S. (2019): Operability Testing of Command, Control, Communications, Computers, and Intelligence (C4I) Systems. In: Handbook of Human Factors Testing and Evaluation: CRC Press, S. 433–456.
- Carmenta. Online verfügbar unter <https://carmenta.com/en/>.
- Cassenti, Daniel N.; Veksler, Vladislav D. (2017): Using cognitive modeling for adaptive automation triggering. In: International Conference on Applied Human Factors and Ergonomics: Springer, S. 378–390.
- Causse, Mickaël; Chua, Zarrin; Peysakhovich, Vsevolod; Del Campo, Natalia; Matton, Nadine (2017): Mental workload and neural efficiency quantified in the prefrontal cortex using fNIRS. In: *Scientific reports* 7 (1), S. 1–15.
- Chaouachi, Maher; Frasson, Claude (2012): Mental workload, engagement and emotions: an exploratory study for intelligent tutoring systems. In: International Conference on Intelligent Tutoring Systems: Springer, S. 65–71.
- Chen, Fang; Zhou, Jianlong; Wang, Yang; Yu, Kun; Arshad, Syed Z.; Khawaji, Ahmad; Conway, Dan (2016): Robust multimodal cognitive load measurement: Springer.
- Childs, Jerry M.; Spears, William D. (1986): Flight-skill decay and recurrent training. In: *Perceptual and motor skills* 62 (1), S. 235–242.
- Cinaz, Burcu; La Marca, Roberto; Arnrich, Bert; Tröster, Gerhard (2010): Monitoring of mental workload levels. In: 97289391.
- Clegg, Chris; Ravden, Susannah; Corbett, Martin; Johnson, Graham (1989): Allocating functions in computer integrated manufacturing: a review and a new method. In: *Behaviour & Information Technology* 8 (3), S. 175–190.
- Cortes, Corinna; Vapnik, Vladimir (1995): Support-vector networks. In: *Machine learning* 20 (3), S. 273–297.
- Cover, Thomas; Hart, Peter (1967): Nearest neighbor pattern classification. In: *IEEE transactions on information theory* 13 (1), S. 21–27.
- Cummings, Mary: Automation bias in intelligent time critical decision support systems. In: AIAA 1st Intelligent Systems Technical Conference, S. 6313.

- Dehais, Frédéric; Causse, Mickaël; Vachon, François; Tremblay, Sébastien (2012): Cognitive conflict in human–automation interactions: a psychophysiological study. In: *Applied ergonomics* 43 (3), S. 588–595.
- Di Gao; Yin, Guisheng; Cheng, Weijie; Feng, Xiaoning (2012): Non-invasive eye tracking technology based on corneal reflex. In: *Procedia Engineering* 29, S. 3608–3612.
- Donath, Diana; Schulte, Axel (2015): Behavior based task and high workload determination of pilots guiding multiple UAVs. In: *Procedia Manufacturing* 3, S. 990–997.
- Dorneich, Michael C.; Ververs, Patricia May; Mathan, Santosh; Whitlow, Stephen; Hayes, Caroline C. (2012): Considering etiquette in the design of an adaptive system. In: *Journal of Cognitive Engineering and Decision Making* 6 (2), S. 243–265.
- Duchowski, Andrew T. (2003): Visual attention. In: *Eye Tracking Methodology: Theory and Practice*: Springer, S. 3–15.
- Dupre, Anthony; Vincent, Sarah; Iaizzo, Paul A. (2005): Basic ECG theory, recordings, and interpretation. In: *Handbook of cardiac anatomy, physiology, and devices*: Springer, S. 191–201.
- Electrophysiology, Task Force of the European Society of Cardiology the North American Society of Pacing (1996): Heart rate variability: standards of measurement, physiological interpretation, and clinical use. In: *Circulation* 93 (5), S. 1043–1065.
- Endsley, Mica R. (1988a): Design and evaluation for situation awareness enhancement. In: *Proceedings of the Human Factors Society annual meeting*: SAGE Publications Sage CA: Los Angeles, CA (32), S. 97–101.
- Endsley, Mica R. (1988b): Situation awareness global assessment technique (SAGAT). In: *Proceedings of the IEEE 1988 national aerospace and electronics conference*: IEEE, S. 789–795.
- Endsley, Mica R. (1996): Automation and situation awareness. In: *Automation and human performance: Theory and applications* 20, S. 163–181.
- Endsley, Mica R. (2001): Designing for situation awareness in complex systems. In: *Proceedings of the Second International Workshop on symbiosis of humans, artifacts and environment*, S. 1–14.
- Endsley, Mica R.; Kiris, Esin O. (1995): The out-of-the-loop performance problem and level of control in automation. In: *Human factors* 37 (2), S. 381–394.
- Faure, Vérane; Lobjois, Régis; Benguigui, Nicolas (2016): The effects of driving environment complexity and dual tasking on drivers' mental workload and eye blink

behavior. In: *Transportation research part F: traffic psychology and behaviour* 40, S. 78–90.

Fernandes, Steven Lawrence; Gurupur, Varadraj Prabhu; Sunder, Nayak Ramesh; Arunkumar, N.; Kadry, Seifedine (2020): A novel nonintrusive decision support approach for heart rate measurement. In: *Pattern Recognition Letters* 139, S. 148–156.

Fortin-Côté, Alexis; Beaudin-Gagnon, Nicolas; Campeau-Lecours, Alexandre; Tremblay, Sébastien; Jackson, Philip L. (2019): Affective Computing Out-of-The-Lab: The Cost of Low Cost. In: 2019 IEEE International Conference on Systems, Man and Cybernetics (SMC): IEEE, S. 4137–4142.

Freund, Yoav; Schapire, Robert; Abe, Naoki (1999): A short introduction to boosting. In: *Journal-Japanese Society For Artificial Intelligence* 14 (771-780), S. 1612.

Friedrich, Elisabeth V. C.; Scherer, Reinhold; Sonnleitner, Kristina; Neuper, Christa (2011): Impact of auditory distraction on user performance in a brain–computer interface driven by different mental tasks. In: *Clinical Neurophysiology* 122 (10), S. 2003–2009.

Friedrich, Maik; Rußwinkel, Nele; Möhlenbrink, Christoph (2017): A guideline for integrating dynamic areas of interests in existing set-up for capturing eye movement: Looking at moving aircraft. In: *Behavior research methods* 49 (3), S. 822–834.

Frohm, Jörgen; Lindström, Veronica; Winroth, Mats; Stahre, Johan (2008): Levels of automation in manufacturing. In: *Ergonomia*.

Fu, Shimin; Parasuraman, Raja (2006): Event-Related Potentials (ERPs) in Neuroergonomics.

Fuchs, Sven; Hale, Kelly S.; Berka, C.; Levendowski, D.; Juhnke, J. (2006): Physiological sensors cannot effectively drive system mitigation alone. In: *Foundations of Augmented Cognition*, S. 193–200.

Gevins, Alan; Smith, Michael E. (2003): Neurophysiological measures of cognitive workload during human–computer interaction. In: *Theoretical Issues in Ergonomics Science* 4 (1-2), S. 113–131.

Gopher, Daniel; Armony, Lilach; Greenshpan, Yaakov (2000): Switching tasks and attention policies. In: *Journal of Experimental Psychology: General* 129 (3), S. 308.

Gordon, Jean; Shortliffe, Edward H. (1984): The Dempster-Shafer theory of evidence. In: *Rule-Based Expert Systems: The MYCIN Experiments of the Stanford Heuristic Programming Project* 3, S. 832–838.

Grier, Rebecca; Wickens, Christopher; Kaber, David; Strayer, David; Boehm-Davis, Deborah; Trafton, J. Gregory; St. John, Mark (2008): The red-line of workload: Theory,

research, and design. In: Proceedings of the human factors and ergonomics society annual meeting: SAGE Publications Sage CA: Los Angeles, CA (52), S. 1204–1208.

Haddadin, Sami; Parusel, Sven; Johannsmeier, Lars; Golz, Saskia; Gabl, Simon; Walch, Florian et al. (2022): The Franka Emika Robot: A Reference Platform for Robotics Research and Education. In: *IEEE Robotics & Automation Magazine*.

Harris Sr, Randall L.; Mixon, Randolph W. (1979): Advanced transport operation effects on pilot scan patterns. In: Proceedings of the Human Factors Society Annual Meeting: SAGE Publications Sage CA: Los Angeles, CA (23), S. 347–351.

Hart, Sandra G. (2006): NASA-task load index (NASA-TLX); 20 years later. In: Proceedings of the human factors and ergonomics society annual meeting: SAGE Publications Sage CA: Los Angeles, CA (50), S. 904–908.

Hart, Sandra G.; Staveland, Lowell E. (1988): Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In: *Advances in psychology*, Bd. 52: Elsevier, S. 139–183.

Hasanzadeh, Sogand; Esmaeili, Behzad; Dodd, Michael D. (2016): Measuring construction workers' real-time situation awareness using mobile eye-tracking. In: *Construction Research Congress 2016*, S. 2894–2904.

Heard, Jamison; Harriott, Caroline E.; Adams, Julie A. (2018): A survey of workload assessment algorithms. In: *IEEE Transactions on Human-Machine Systems* 48 (5), S. 434–451.

Heilemann, Felix; Hollatz, Felix (2020): Integration of mental resources in the planning of manned-unmanned teaming missions: concept, implementation and evaluation. In: *AIAA Scitech 2020 Forum*, S. 1111.

Heilemann, Felix; Schulte, Axel (2020): Experimental evaluation of an adaptive planning assistance system in manned unmanned teaming missions. In: *International Conference on Human-Computer Interaction*: Springer, S. 371–382.

Herff, Christian; Heger, Dominic; Fortmann, Ole; Hennrich, Johannes; Putze, Felix; Schultz, Tanja (2014): Mental workload during n-back task—quantified in the prefrontal cortex using fNIRS. In: *Frontiers in human neuroscience* 7, S. 935.

Hoffman, Robert R.; Feltovich, Paul J.; Ford, Kenneth M.; Woods, David D. (2002): A rose by any other name... would probably be given an acronym [cognitive systems engineering]. In: *IEEE Intelligent Systems* 17 (4), S. 72–80.

Hollnagel, Erik (1993): Human reliability analysis. In: *Context and control*.

- Hollnagel, Erik (1998): Cognitive reliability and error analysis method (CREAM): Elsevier.
- Hollnagel, Erik; Woods, David D. (2005): Joint cognitive systems. Foundations of cognitive systems engineering. Boca Raton FL: Taylor & Francis.
- Honecker, Fabian; Schulte, Axel (2017): Automated online determination of pilot activity under uncertainty by using evidential reasoning. In: International Conference on Engineering Psychology and Cognitive Ergonomics: Springer, S. 231–250.
- Horrey, William J.; Wickens, Christopher D. (2001): Supporting battlefield situation assessment through attention guidance and diagnostic aiding: A cost-benefit and depth of processing analysis. ILLINOIS UNIV AT URBANA-CHAMPAIGN SAVOYAVIATION RESEARCH LAB.
- Inagaki; T: Adaptive Automation: Sharing and Trading of Control.
- Johannsen, Gunnar (2013): Mensch-Maschine-Systeme: Springer-Verlag.
- Jones, Debra G.; Endsley, Mica R. (2004): Use of real-time probes for measuring situation awareness. In: *The International Journal of Aviation Psychology* 14 (4), S. 343–367.
- Kaber, David B.; Endsley, Mica R. (2004): The effects of level of automation and adaptive automation on human performance, situation awareness and workload in a dynamic control task. In: *Theoretical Issues in Ergonomics Science* 5 (2), S. 113–153.
- Kahneman, Daniel (1973): Attention and effort: Citeseer (1063).
- Kerstholt, Jose h.; Passenier, Peter O. (2000): Fault management in supervisory control: the effect of false alarms and support. In: *Ergonomics* 43 (9), S. 1371–1389.
- Kessel, Colin J.; Wickens, Christopher D. (1982): The transfer of failure-detection skills between monitoring and controlling dynamic systems. In: *Human factors* 24 (1), S. 49–60.
- Kotsiantis, Sotiris B.; Zaharakis, I.; Pintelas, P. (2007): Supervised machine learning: A review of classification techniques. In: *Emerging artificial intelligence applications in computer engineering* 160 (1), S. 3–24.
- KUBIOS OY. Online verfügbar unter <https://www.kubios.com/>.
- Laminar Research. Online verfügbar unter <https://www.x-plane.com/>.
- Lee, John D.; See, Katrina A. (2004): Trust in automation: Designing for appropriate reliance. In: *Human factors* 46 (1), S. 50–80.
- Li, Kai; Rüdiger, Heinz; Ziemssen, Tjalf (2019): Spectral analysis of heart rate variability: time window matters. In: *Frontiers in neurology* 10, S. 545.

Lienhart, Rainer; Maydt, Jochen (2002): An extended set of haar-like features for rapid object detection. In: Proceedings. international conference on image processing: IEEE (1), S. I.

Lim, JungKyu Rhys; Liu, Brooke Fisher; Egnoto, Michael (2019): Cry wolf effect? Evaluating the impact of false alarms on public responses to tornado alerts in the southeastern United States. In: *Weather, climate, and society* 11 (3), S. 549–563.

Lindner, Sebastian; Mund, Dennis; Schulte, Axel (2022): How Human-Autonomy Teams change the Role of future Fighter Pilots: An Experimental Assessment. In: AIAA SCITECH 2022 Forum, S. 2551.

Lindner, Sebastian; Schulte, Axel (2020): Human-in-the-loop evaluation of a manned-unmanned system approach to derive operational requirements for military air missions. In: International Conference on Human-Computer Interaction: Springer, S. 341–356.

Lotz, Alexander; Russwinkel, Nele; Wagner, Thomas; Wohlfarth, Enrico (2020): An adaptive assistance system for subjective critical driving situations: understanding the relationship between subjective and objective complexity.

Lyu, Nengchao; Xie, Lian; Wu, Chaozhong; Fu, Qiang; Deng, Chao (2017): Driver's cognitive workload and driving performance under traffic sign information exposure in complex environments: A case study of the highways in China. In: *International journal of environmental research and public health* 14 (2), S. 203.

Maiwald, Felix; Schulte, Axel (2016): Enhancing military helicopter pilot assistant systems through resource adaptive dialogue management. In: *Advances in Aviation Psychology. Ashgate Studies in Human Factors and Flight Operations*, S. 177–196.

Maiwald, Johannes Felix Matthias (2015): Maschinelle Beanspruchungsprädiktion zur ressourcengerechten Adaption eines Pilotenassistenzsystems: Kovač.

Mark, Gloria; Gudith, Daniela; Klocke, Ulrich (2008): The cost of interrupted work: more speed and stress. In: Proceedings of the SIGCHI conference on Human Factors in Computing Systems, S. 107–110.

Marquart, Gerhard; Cabrall, Christopher; Winter, Joost de (2015): Review of eye-related measures of drivers' mental workload. In: *Procedia Manufacturing* 3, S. 2854–2861.

Mathôt, Sebastiaan (2018): Pupillometry: psychology, physiology, and function. In: *Journal of Cognition* 1 (1).

Matthews, Gerald; Davies, D. Roy (2001): Individual differences in energetic arousal and sustained attention: A dual-task study. In: *Personality and individual Differences* 31 (4), S. 575–589.

- McCracken, J. H.; Aldrich, T. B. (1984): Analyses of selected LHX mission functions: Implications for operator workload and system automation goals. ANACAPA SCIENCES INC FORT RUCKER AL.
- McFarlane, Daniel C.; Latorella, Kara A. (2002): The scope and importance of human interruption in human-computer interaction design. In: *Human-Computer Interaction* 17 (1), S. 1–61.
- Mehler, Bruce; Reimer, Bryan; Wang, Ying (2011): A comparison of heart rate and heart rate variability indices in distinguishing single-task driving and driving under secondary cognitive workload. In: Proceedings of the sixth international driving symposium on human factors in driver assessment, training and vehicle design, S. 590–597.
- Mitchell, Diane K. (2000): Mental workload and ARL workload modeling tools. Army Research Lab Aberdeen Proving Ground MD.
- Moray, Neville: Human Information Processing and Supervisory Control. Massachusetts Inst of Tech Cambridge Man-Machine Systems Lab.
- Mosier, Kathleen L.; Skitka, Linda J.; Heers, Susan; Burdick, Mark (1998): Automation bias: Decision making and performance in high-tech cockpits. In: *The International Journal of Aviation Psychology* 8 (1), S. 47–63.
- Mukherjee, Shalini; Yadav, Rajeev; Yung, Iris; Zajdel, Daniel P.; Oken, Barry S. (2011): Sensitivity to mental effort and test–retest reliability of heart rate variability measures in healthy seniors. In: *Clinical Neurophysiology* 122 (10), S. 2059–2066.
- Mund, D.; Schulte, A. (2018): Model-and observation-based workload assessment and activity determination in manned-unmanned teaming missions. In: 33rd EAAP Conference (European Association for Aviation Psychology): Dubrovnik, Croatia, S. 24–28.
- Mund, Dennis; Heilemann, Felix; Reich, Florian; Denk, Elisabeth; Donath, Diana; Schulte, Axel (2017): Experimental Analysis of Behavioral Workload Indicators to Facilitate Adaptive Automation for Fighter-UCAV Interoperability. In: *Advances in Human Factors in Robots and Unmanned Systems*: Springer, S. 239–250.
- Mund, Dennis; Pavlidis, Evgeni; Masters, Matthew; Schulte, Axel (2020): A conceptual augmentation of a pilot assistant system with physiological measures. In: *International Conference on Intelligent Human Systems Integration*: Springer, S. 959–965.
- Muslim, H.; Itoh, M. (2019): Trust and acceptance of adaptive and conventional collision avoidance systems. In: *IFAC-PapersOnLine* 52 (19), S. 55–60.

Nassef, Ashraf; Mahfouf, Mahdi; Ting, Ching-Hua; El-Samahy, Emad; Linkens, Derek A.; Denaï, Mouloud Azzedine (2010): Hybrid Physiological Modeling of Subjects Undergoing Cyclic Physical Loading. In: *Biosignals*, S. 252–257.

Neerincx, Mark A. (2003): Cognitive task load analysis: allocating tasks and designing support. In: *Handbook of cognitive task design 2003*, S. 283–305.

Neerincx, Mark A.; van Besouw, N. J.P. (2001): Cognitive task load: a function of time occupied, level of information processing and task-set switches.

Nitschke, Christian; Nakazawa, Atsushi; Takemura, Haruo (2013): Corneal imaging revisited: An overview of corneal reflection analysis and applications. In: *IPSJ Transactions on Computer Vision and Applications 5*, S. 1–18.

North, Robert A.; Riley, Victor A. (1989): W/INDEX: A predictive model of operator workload. In: *Applications of human performance models to system design*: Springer, S. 81–89.

Noy, Ian Y.; Shinar, David; Horrey, William J. (2018): Automated driving: Safety blind spots. In: *Safety Science* 102, S. 68–78.

Pan, Jiapu; Tompkins, Willis J. (1985): A real-time QRS detection algorithm. In: *IEEE transactions on biomedical engineering* (3), S. 230–236.

Parasuraman, Raja; Cosenzo, Keryl A.; Visser, Ewart de (2009): Adaptive automation for human supervision of multiple uninhabited vehicles: Effects on change detection, situation awareness, and mental workload. In: *Military Psychology* 21 (2), S. 270–297.

Parasuraman, Raja; Manzey, Dietrich H. (2010): Complacency and bias in human use of automation: An attentional integration. In: *Human factors* 52 (3), S. 381–410.

Parasuraman, Raja; Molloy, Robert; Singh, Indramani L. (1993): Performance consequences of automation-induced 'complacency'. In: *The International Journal of Aviation Psychology* 3 (1), S. 1–23.

Parasuraman, Raja; Mouloua, Mustapha; Molloy, Robert (1996): Effects of adaptive task allocation on monitoring of automated systems. In: *Human factors* 38 (4), S. 665–679.

Parasuraman, Raja; Riley, Victor (1997): Humans and automation: Use, misuse, disuse, abuse. In: *Human factors* 39 (2), S. 230–253.

Parasuraman, Raja; Rizzo, Matthew (2006): *Neuroergonomics: The brain at work*: Oxford University Press (3).

Parasuraman, Raja; Sheridan, Thomas B.; Wickens, Christopher D. (2000): A model for types and levels of human interaction with automation. In: *IEEE Transactions on systems, man, and cybernetics-Part A: Systems and Humans* 30 (3), S. 286–297.

Patriarca, Riccardo; Di Gravio, Giulio; Woltjer, Rogier; Costantino, Francesco; Praetorius, Gesa; Ferreira, Pedro; Hollnagel, Erik (2020): Framing the FRAM: A literature review on the functional resonance analysis method. In: *Safety Science* 129, S. 104827.

Prinzel III, Lawrence J.; Kaber, David B. (2006): Adaptive and Adaptable Automation Design: A Critical Review of the Literature and Recommendations for Future Research.

Ramaraj, Preeti; Klenk, Matt; Mohan, Shiwali (Hg.) (2020): Understanding Intentions in Human Teaching to Design Interactive Task Learning Robots.

Rasmussen, Jens (1983): Skills, rules, and knowledge; signals, signs, and symbols, and other distinctions in human performance models. In: *IEEE transactions on systems, man, and cybernetics* (3), S. 257–266.

Reiser Simulation and Training GmbH. Online verfügbar unter <http://reiser-systemtechnik.info/>.

Ridella, Sandro; Rovetta, Stefano; Zunino, Rodolfo (1997): Circular backpropagation networks for classification. In: *IEEE transactions on neural networks* 8 (1), S. 84–97.

Roy, Raphaëlle N.; Bonnet, Stéphane; Charbonnier, Sylvie; Campagne, Aurélie (2016): Efficient workload classification based on ignored auditory probes: a proof of concept. In: *Frontiers in human neuroscience* 10, S. 519.

Rusnock, Christina F.; Geiger, Christopher D. (2016): Designing adaptive systems: selecting an invoking threshold to improve human performance. In: *International journal of human factors and ergonomics* 4 (3-4), S. 292–315.

Satchell, Paul M. (2016): Cockpit monitoring and alerting systems: Routledge.

Sauer, Juergen; Chavaillaz, Alain; Wastell, David (2017): On the effectiveness of performance-based adaptive automation. In: *Theoretical Issues in Ergonomics Science* 18 (3), S. 279–297.

Scerbo, Mark W. (2001): The efficacy of psychophysiological measures for implementing adaptive technology: NASA Langley Research Center.

Schapiro, Robert E. (2013): Explaining adaboost. In: *Empirical inference*: Springer, S. 37–52.

Schmidt, Gerald J.; Rittger, Lena (2016): Guiding driver visual attention with LEDs. In: *Proceedings of the 9th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, S. 279–286.

- Schmitt, Fabian; Brand, Yannick; Rudnick, Georg; Schulte, Axel (Hg.) (2018): Experimental evaluation of a cooperative automation approach for manned-unmanned teaming in future military helicopter missions.
- Schmorrow, Dylan; Stanney, Kay M.; Wilson, Glenn; Young, Peter (2006): Augmented cognition in human–system interaction. In: *Handbook of human factors and ergonomics*, S. 1364–1383.
- Schreiber, Günter; Stemmer, Andreas; Bischoff, Rainer (2010): The fast research interface for the kuka lightweight robot. In: IEEE workshop on innovative robot control architectures for demanding (Research) applications how to modify and enhance commercial controllers (ICRA 2010): Citeseer, S. 15–21.
- SCHUCK, MARY MARGARET (1996): Development of equal-interval task rating scales and task conflict matrices as predictors of attentional demand. In: *Ergonomics* 39 (3), S. 345–357.
- Schulte, Axel; Donath, Diana (2011): Measuring self-adaptive UAV operators' load-shedding strategies under high workload. In: International Conference on Engineering Psychology and Cognitive Ergonomics: Springer, S. 342–351.
- Schwarz, Jessica; Fuchs, Sven (2018): Validating a "Real-Time Assessment of Multidimensional User State"(RASMUS) for Adaptive Human-Computer Interaction. In: 2018 IEEE International Conference on Systems, Man, and Cybernetics (SMC): IEEE, S. 704–709.
- Schwarz, Jessica Carolin; Erstgutachter, Vorgeschlagener; Rinkenauer, P. GerhardD; Zweitgutachter, Vorgeschlagener; Krems, Josef F. (2019): Multifaktorielle Echtzeitdiagnose des Nutzerzustands in adaptiver Mensch-Maschine-Interaktion.
- Schwerd, Simon; Schulte, Axel (2020): Experimental Validation of an Eye-Tracking-Based Computational Method for Continuous Situation Awareness Assessment in an Aircraft Cockpit. In: International Conference on Human-Computer Interaction: Springer, S. 412–425.
- Sheridan, Thomas B. (1992): Telerobotics, automation, and human supervisory control: MIT press.
- Sheridan, Thomas B.; Verplank, William L. (1978): Human and computer control of undersea teleoperators. Massachusetts Inst of Tech Cambridge Man-Machine Systems Lab.
- Simons, Daniel J.; Chabris, Christopher F. (1999): Gorillas in our midst: Sustained inattention blindness for dynamic events. In: *perception* 28 (9), S. 1059–1074.
- Singh, Gaganpreet (2021): Enhancing manned unmanned teaming with physiological computing and adaptive interaction: Toulouse, ISAE.

- Smart Eye AB. Online verfügbar unter <https://smarteye.se/>.
- Son, Joonwoo; Park, Suwan (2011): Cognitive workload estimation through lateral driving performance. In: Proceedings of the 16th Asia Pacific Automotive Engineering Conference, S. 6–8.
- Sorkin, Robert D. (1988): Why are people turning off our alarms? In: *The Journal of the Acoustical Society of America* 84 (3), S. 1107–1108.
- Sperandio, J. C. (1971): Variation of operator's strategies and regulating effects on workload. In: *Ergonomics* 14 (5), S. 571–577.
- Spread Concepts LLC. Online verfügbar unter <http://www.spread.org/>.
- Staal, Mark A. (2004): Stress, cognition, and human performance: A literature review and conceptual framework: Citeseer.
- Stroop, J. Ridley (1935): Studies of interference in serial verbal reactions. In: *Journal of experimental psychology* 18 (6), S. 643.
- Stuiber, Gitta (2006): Studie zur Untersuchung der Auswirkungen von Koffein auf den Pupillographischen Schläfrigkeitstest bei gesunden Probanden: Universität Tübingen.
- Taatgen, Niels A.; Lebiere, Christian; Anderson, John R. (2006): Modeling paradigms in ACT-R. In: *Cognition and multi-agent interaction: From cognitive modeling to social simulation*, S. 29–52.
- Taatgen, Niels A.; Lee, Frank J. (2003): Production compilation: A simple mechanism to model complex skill acquisition. In: *Human factors* 45 (1), S. 61–76.
- Taylor, Richard M. (2017): Situational awareness rating technique (SART): The development of a tool for aircrew systems design. In: *Situational awareness*: Routledge, S. 111–128.
- Theißing, Nikolaus; Schulte, Axel (2016): Designing a support system to mitigate pilot error while minimizing out-of-the-loop-effects. In: *International Conference on Engineering Psychology and Cognitive Ergonomics*: Springer, S. 439–451.
- U.S. Department of Transportation (1972): Lockheed Model L-1011 Eastern Airlines Flight 401, N310EA. U.S. Department of Transportation. Online verfügbar unter https://lessonslearned.faa.gov/ll_main.cfm?TabID=1&LLID=8.
- Veltman, J. A.; Jansen, C. (2005): The role of operator state assessment in adaptive automation.

- Visser, Ewart J. de; Cohen, Marvin; Freedy, Amos; Parasuraman, Raja: A design methodology for trust cue calibration in cognitive agents. In: International conference on virtual, augmented and mixed reality: Springer, S. 251–262.
- Volz, Katherine; Yang, Euijung; Dudley, Rachel; Lynch, Elizabeth; Dropps, Maria; Dorneich, Michael C. (2016): An evaluation of cognitive skill degradation in information automation. In: Proceedings of the Human Factors and Ergonomics Society Annual Meeting: SAGE Publications Sage CA: Los Angeles, CA (60), S. 191–195.
- Volz, Katherine M.; Dorneich, Michael C. (2020): Evaluation of Cognitive Skill Degradation in Flight Planning. In: *Journal of Cognitive Engineering and Decision Making* 14 (4), S. 263–287. DOI: 10.1177/1555343420962897.
- Vuksanović, Vesna; Gal, Vera (2007): Heart rate variability in mental stress aloud. In: *Medical engineering & physics* 29 (3), S. 344–349.
- Wang, Weihong; Li, Zhidong; Wang, Yang; Chen, Fang (2013): Indexing cognitive workload based on pupillary response under luminance and emotional changes. In: Proceedings of the 2013 international conference on Intelligent user interfaces, S. 247–256.
- Wickens, Christopher D.: Attentional tunneling and task management. In: 2005 International Symposium on Aviation Psychology, S. 812.
- Wickens, Christopher D. (2002): Multiple resources and performance prediction. In: *Theoretical Issues in Ergonomics Science* 3 (2), S. 159–177.
- Wickens, Christopher D.; Rice, Stephen; Keller, David; Hutchins, Shaun; Hughes, Jamie; Clayton, Krisstal (2009): False alerts in air traffic control conflict alerting system: Is there a “cry wolf” effect? In: *Human factors* 51 (4), S. 446–462.
- Winter, J. C. F. de; Dodou, D. (2014): Why the Fitts list has persisted throughout the history of function allocation. In: *Cogn Tech Work* 16 (1), S. 1–11. DOI: 10.1007/s10111-011-0188-1.
- Wong, Tzu-Tsung (2015): Performance evaluation of classification algorithms by k-fold and leave-one-out cross validation. In: *Pattern Recognition* 48 (9), S. 2839–2846.
- Xu, Jie; Wang, Yang; Chen, Fang; Choi, Eric (2011): Pupillary response based cognitive workload measurement under luminance changes. In: IFIP Conference on Human-Computer Interaction: Springer, S. 178–185.
- Xu, Jing; Kendrick, Karen; Bowers, Alex R. (2022): Clinical report: experiences of a driver with vision impairment when using a Tesla car. In: *Optometry and vision science* 99 (4), S. 417–421.

Yerkes, Robert M.; Dodson, John D. (1908): The relation of strength of stimulus to rapidity of habit-formation. In: *Punishment: Issues and experiments*, S. 27–41.