Driver Monitoring mit Fokus auf Vigilanz

Fahrerzustandserkennung durch Fusion von physiologischen Daten

DISSERTATION

eingereicht zur Erlangung
des akademischen Grades

Doktor-Ingenieur

an der

Fakultät für Elektrotechnik und Informationstechnik
Technische Universität Dortmund

von

Alexander Hugenroth, M.Sc.

Tag der Einreichung: 27. Mai 2024

Erster Gutachter: Univ.-Prof. Dr.-Ing. Prof. h.c. Dr. h.c. Torsten Bertram
Zweiter Gutachter: Prof. Dr. Sebastian Pannasch

Tag des Kolloquiums: 18. Dezember 2024

Danksagung

An erster Stelle möchte ich mich bei Herrn Professor Dr.-Ing. Prof. h.c. Dr. h.c. Torsten Bertram als wissenschaftlicher Betreuer dieser Arbeit bedanken. Ebenso freue ich mich über die Unterstützung von Herrn Professor Dr. Pannasch für die Unterstützung als Zweitprüfer.
Da die Arbeit in Zusammenarbeit mit der Firma HELLA GmbH & Co. KGaA (nun unter der Dachmarke Forvia) entstanden ist, möchte ich mich für die Unterstützung und Möglichkeit zur freien Ausübung der Forschung bedanken. Mein Dank gilt dabei allen ehemaligen und aktuellen Kollegen des E-LABs, besonders Dr. Yuliang Sun und Dr. Daniel Brieske als jahrelange Bürokollegen. Ebenso erhielt ich große Unterstützung von dem Laboringenieur Tobias Brandt, der mir in allen Fragen rund um den Laborbetrieb und dem Versuchsaufbau zur Seite stand. Dazu möchte ich die richtungsweisenden fachlichen Diskussionen und den netten sowie kollegialen Umgang ausgehend von der Leitung des E-LABs (ehemals) Dr. Alexander Warnicke und Dr. Claas Tebrügge honorieren.
Auch bedanken möchte ich mich bei allen Mitarbeiterinnen und Mitarbeitern, Gastwissenschaftlerinnen und Gastwissenschaftlern sowie externen Doktorandinnen und Doktoranden des Lehrstuhls für Regelungssystemtechnik der TU Dortmund. Bei den Kaffepausen hat regelmäßig ein angeregter Austausch stattgefunden und die Teamevents waren stets ein Highlight im Jahr.
Ebenso haben zur Umsetzung, der vorliegenden Dissertation, studentische Arbeiten beigetragen. Hiermit möchte ich den beteiligten Studierenden meinen Dank ausdrücken.
Mein besonderer Dank gilt meiner ganzen Familie. Die liebevolle und ausdauernde Unterstützung meiner Eltern Monika und Norbert Hugenroth ist dabei hervorzuheben.

Kurzfassung

Das automatisierte Fahren entwickelt sich stetig weiter. Fahrerassistenzsysteme übernehmen immer mehr Aufgaben des Fahrers oder der Fahrerin. Die Längs- und Querdynamik, also das Beschleunigen, Abbremsen und Lenken, kann hierbei vom Fahrzeug übernommen werden. Für die meisten Systeme besteht eine Notwendigkeit, dass der Fahrer oder die Fahrerin zu jedem Zeitpunkt aufmerksam bleibt. Das Verhalten des Systems muss vom Fahrer oder Fahrerin jederzeit überprüft werden, sodass die komplette Fahraufgabe in Kürze übernommen werden kann. Folglich gewinnt die Fahrerzustandserkennung an Relevanz, um diese Bedingung abzusichern. Bisher sind in diesem Feld die Müdigkeitserkennung sowie Verfahren zur Auswertung der Blickrichtung des Fahrers oder der Fahrerin in Serie.
Der Fokus der vorliegenden Arbeit liegt hingegen auf den psychologischen Effekten, die über die gerade genannten Systeme hinausgehen. Unter den anfangs beschriebenen veränderten Voraussetzungen wird die Betrachtung der Daueraufmerksamkeit interessant. Die Vigilanz beschäftigt sich mit der Langzeitaufmerksamkeit bei nur seltenen und schwachen Reizen. Um diesen Zustand aufzunehmen, werden in der vorliegenden Arbeit physiologische Merkmale (von Herz, Augen und Körperbewegungen) mit verschiedenen Sensoren während einer automatisierten Fahrt im Fahrsimulator bei einer Gruppe von Probandinnen und Probanden aufgezeichnet. Da zwischen den physiologischen Merkmalen und diesem speziellen psychologischen Zustand (der Vigilanz) kein analytischer Zusammenhang bekannt ist, werden verschiedene Modelle auf Basis von selbstlernenden, künstlichen neuronalen Netzwerken im Rahmen der vorliegenden Arbeit entwickelt. Diese werden mit Daten aus psychologischen Probandenstudien trainiert und ausgewertet. Eine wesentliche Differenz zwischen den Modellen besteht in der Art der Datenfusion. Es wird sowohl ein Modell auf Basis von Fusion auf Merkmalsebene als auch ein Modell auf Basis von Fusion auf Entscheidungsebene evaluiert. Mit den unterschiedlichen Testdatensätzen zeigen beide Modelle eine ähnliche Performanz. Bei der Betrachtung der Ausfallsicherheit zeigt sich jedoch ein Vorteil in der Fusion auf Entscheidungsebene. So wird dieses Modell für den Einsatz als Aufmerksamkeitserkennung vorgeschlagen und könnte in einem automatisierten Fahrzeug als Warnsystem für erhöhte Sicherheit sorgen. Hiermit könnten nach der Erkennung von Unaufmerksamkeit verschiedene Maßnahmen eingeleitet werden, um den Fahrer wieder in einen anderen Zustand der Aufmerksamkeit zu versetzen.

Inhaltsverzeichnis

Nomenklatur ii

1. Einleitung 1

2. Psychologie und Physiologie bei längerer automatisierter Fahrt 5

3. Datenverarbeitung und Fusion mittels künstlichen neuronalen Netzwerken 18

4. Machbarkeitsstudie zum Aufmerksamkeitsverhalten 25
4.1. Design einer Signalentdeckungsaufgabe und Aufbau der Studie 25
4.2. Ergebnisse der Fragebögen und Reaktionszeiten 29

5. Probanden-Studie zur Vigilanzuntersuchung im Fahrsimulator 33
5.1. Design einer Vigilanzaufgabe mit unterschiedlichen Reizen 33
5.2. Psychologische Ergebnisse der Studie 38

6. Datenverarbeitung und Fusion von verschiedenen physiologischen Merkmalen 46
6.1. Datenverarbeitung zur Erstellung von Merkmalen aus physiologischen Daten 46
6.2. Fusion der physiologischen Daten auf Merkmalsebene 50
6.3. Fusion der physiologischen Daten auf Entscheidungsebene 55

7. Modell zur Klassifikation von Fahreraufmerksamkeit mit Fokus auf Vigilanz 62
7.1. Performanz der verschiedenen Fusionsmethoden 62
7.2. Modell zur Bestimmung des Aufmerksamkeitsverhaltens 86

8. Zusammenfassung und Ausblick 90

A. Anhang 100
A.1. Messsystem zur Aufnahme von physiologischen Daten 100
A.1.1. Sensoraufbau 100
A.2. Vigilanzstudie 103

Literatur 117

Nomenklatur

F	fehlerhafter Datensatz
G	Genauigkeit
H	Herzraten-Variabilität
HF	Herzfrequenz
N	Anzahl einer Menge
P	Präzision
R	Aussfallquote
S	Spezifität
T	Testdatensatz
V	negativen Vorhersagewert
Z_{H}	Zeitspanne der Herzraten-Variabilität
$\mathbf{a}$	Ausgangsgatter
$\mathbf{b}$	Gewichte für den Bias
$\mathbf{c}$	Zellstatus
$\mathbf{e}$	Eingangsgatter
$\mathbf{h}$	Gewichte für den versteckten Status
$\mathbf{h}$	versteckter Status
$\mathbf{v}$	Vergessensgatter
$\mathbf{w}$	Gewichte für den Merkmalsvektor
$\mathbf{x}$	Merkmalsvektor
$\mathbf{z}$	Zellkandidat
i	Zählindex
n_{fn}	Anzahl falsch-negativer Klassifikationen
n_{fp}	Anzahl falsch-positiver Klassifikationen
n_{ges}	Anzahl aller Klassifikationen
n_{tp}	Anzahl richtig-positiver Klassifikationen
r	Zeiten zwischen zwei Herzschlägen
t	Zeitpunkt
$t+1$	Zeitpunkt nach t
$t-1$	Zeitpunkt vor t
t_{m}	Zeitpunkt der Vorhersage des Modells
t_{tp}	Zeitpunkt der wahren Änderung
t_{v}	Verzugszeit

Abkürzungen und Akronyme

ABS	**A**nti**b**lockier**s**ystem
bpm	englisch: **b**eats **p**er **m**inute
DSSQ	**D**undee **S**tress **S**tate **Q**uestionnaire
EEG	**E**lektro**e**nzephalo**g**ramm

EEI	**E**EG-**E**ngagement-**I**ndex
EKG	**E**lektro**k**ardio**g**ramm
ESP	**e**lektronisches **S**tabilisierungs**p**rogramm
FC	englisch: **f**ully **c**onnected
ID	**Id**entifikationsnummer
KSS	englisch: **K**arolinska **S**leepness **S**cale
LSTM	englisch: **L**ong **S**hort-**T**erm **M**emory
MLP	englisch: **m**ulti **l**ayer **p**erceptron
NIRS	**N**ah**i**nfra**r**ot**s**pektroskopie
PERCLOS	englisch: **per**centage of eye **clos**ure (Merkmal)
SP	englisch: **s**eat **p**reasure (Merkmal)
SSSQ	**S**hort **S**tress **S**tate **Q**uestionnaire
STD	**St**andar**d**abweichung

Griechische Symbole

α	Alpha-Frequenzband des EEG [7,5-13 Hz]
β	Beta-Frequenzband des EEG [13-30 Hz]
δ	Delta-Frequenzband des EEG [1-4 Hz]
γ	Gamma-Frequenzband des EEG [30-44 Hz]
σ_g	Aktivierungsfunktion der Gatter
σ_z	Aktivierungsfunktion des Zellstatus
θ	Theta-Frequenzband des EEG [4-8 Hz]

Mathematische Operatoren

$\circ$	Hadamard-Produkt, elementweise Multiplikation von Vektoren

1 Einleitung

Schon seit mehreren Jahren ist das automatisierte Fahren ein wesentlicher Trend im Automobilbereich, denn mit steigender Automatisierung erhöht sich sowohl der Komfort als auch die Sicherheit der Fahrzeuge. Begonnen hat die Entwicklung mit einfachen Sicherheitssystemen wie das Antiblockiersystem (ABS) und das elektronische Stabilisierungsprogramm (ESP), welche durch Eingriffe in Bremsmanöver die Stabilität des Fahrverhaltens erhöhen. Weitere Aspekte schließen die Umfeld-Erfassung mit ein, wie Parkassistenten, die Parkmöglichkeiten bewerten und die Querführung des Fahrzeugs komplett übernehmen können. Die aktuell neusten Entwicklungen beziehen sich beispielsweise auf Fahrerassistenzsysteme, welche in Stau-Situationen bei niedrigen Geschwindigkeiten über einen Abstandsregel-Tempomaten und Lenkassistenten bereits in der Lage sind sowohl die Längs- als auch die Querführung des Fahrzeugs zu übernehmen. Die verschiedenen Automatisierungen und die zugehörigen Verantwortlichkeiten für den Fahrzeugführenden sowie auch für das Fahrzeug werden in dem SAE-Standard J3016:2021 in Levels eingeteilt. Mit der steigenden Automatisierung sind in diesem Standard die Level 0 bis 2 der Assistenzsysteme und des partiellen automatisierten Fahrens bereits erreicht worden. Bei allen bisher in der Serie eingeführten Systemen gibt es dabei einen Zusammenhang: Der Fahrzeugführer bzw. die Fahrzeugführerin muss jederzeit aufmerksam und permanent in der Lage sein, die Kontrolle des Fahrzeugs in wenigen Sekunden zu übernehmen. Zudem ist der Fahrzeugführer oder die Fahrzeugführerin dazu in der Lage, das Verhalten des Systems zu jeder Zeit zu übersteuern. Dennoch zeichnet sich zur Sicherheits- und Komfortoptimierung ein Trend zu höherer Automatisierung ab, welcher auch Einfluss auf die psychologische Situation beim Fahren haben kann. Dabei ist immer wieder die Rede von Level 2+ Systemen, die sowohl Längs- als auch Querführung des Fahrzeugs übernehmen können. Dies führt zu einem deutlichen Komfortgewinn, aber immer mit dem Hintergrund, dass die Fahrerin oder der Fahrer noch aufmerksam die Automatisierung überprüfen muss. Diese Systeme funktionieren, im Gegensatz zu aktuell zugelassenen Level 3 Systemen, unter wesentlich weniger Voraussetzungen und auch bei höheren Geschwindigkeiten. Diese Systeme sind besonders kritisch für die Aufmerksamkeit, da viele Aufgaben für den Komfortgewinn von dem Fahrzeug übernommen und so die Fahraufgabe zur Unterforderung werden kann.
In diesem Kontext ist der Terminus **Vigilanz** entscheidend, welcher folgendermaßen

für die vorliegende Arbeit definiert wird: Die Vigilanz beschreibt eine Art Daueraufmerksamkeit, welche im Zusammenhang mit einer längeren Aufgabe steht. Dabei wird eine gewisse Reaktionsbereitschaft auf schwache und seltene Reize gefordert. Diese Aspekte finden sich vermehrt bei hoch automatisierten Aufgaben wieder, wie es die Fahrt in einem automatisierten Fahrzeug werden könnte. Eine differenziertere Definition zur Vigilanz folgt im Kapitel 2.
Die vorliegende Arbeit behandelt die Erfassung der Aufmerksamkeit durch verschiedene Fahrermodelle. Hierbei soll das Fahrzeug feststellen können, ob der Fahrer oder die Fahrerin noch laut der gerade genannten Definition aufmerksam ist oder nicht. Diese Erfassung wird technisch über die Messung von unterschiedlichen physiologischen Merkmalen durchgeführt, die dann durch selbstlernende künstliche neuronale Netzwerke eingeordnet werden. Diese Erfassung soll vor und unabhängig von einer Übernahmesituation stattfinden. Ein Übernahmevorgang könnte dann ausgelöst werden und der Umgang mit dem Fahrer oder der Fahrerin verbessert werden.
Da längere automatisierte Fahrten noch nicht verbreitet möglich sind, ist eine Müdigkeitserkennung in vielen Serienfahrzeugen der aktuelle technische Standard. Der Unterschied zwischen Müdigkeit und Vigilanz wird im folgenden Kapitel genauer definiert. An dieser Stelle soll festgehalten werden, dass Müdigkeit oft mit einer weniger genauen Definition als Ursache für einen Aufmerksamkeitsverlust herangezogen wird. Die Müdigkeit hat also sowohl Einfluss auf die Vigilanz als auch auf andere aufmerksamkeitsbezogene Aufgaben wie das manuelle Fahren. Zunächst werden die bisherigen Fahrerzustandssysteme näher beschrieben, um einen Überblick über den Stand der Technik in den heutigen Fahrzeugen zu erlangen.
Die meisten Systeme in Serie benutzen die von [Kra+09] vorgestellte fahrzeugbasierte Messmethode mithilfe von Lenkradbewegungen, um die Müdigkeit zu bestimmen. Da bei einem automatisierten Fahrzeug das Festhalten des Lenkrads nicht mehr notwendig ist, muss dieses Problem auf eine andere Weise gelöst werden.
Von [Sah+12a] werden die Möglichkeiten zur Bestimmung der Müdigkeit bei einem Fahrversuch im Fahrsimulator in vier Gruppen unterteilt: subjektive Messung, fahrzeugbasierte Messungen, Messung des Verhaltens und physiologische Messungen. Die subjektive Messung meint dabei die Erfassung der Müdigkeit durch Fragebögen vor, nach oder während der Fahrt. Bei der fahrzeugbasierten Messung wird die Müdigkeit durch Eingaben des Fahrers oder der Fahrerin an dem Fahrzeug (wie Lenk- und Pedalbewegungen) bestimmt. Die Messung des Verhaltens beschreibt die Erfassung der Müdigkeit durch Bewegungen, Laute oder Emotionen. Unter der physiologischen Messung ist die Erfassung der Müdigkeit über die Aufnahme von Körperfunktionen (wie Herzschlag, Liedschluss, etc.) zu verstehen. Als nächstes werden Literaturbeispiele zu den verschiedenen Messmethoden vorgestellt.
So analysiert [Fel+18a] eine Möglichkeit, die Müdigkeit nur anhand von Kameradaten, also physiologischen Messmethoden, feststellen zu können. Da bei der Erfassung mit einer Kamera die Lichteinflüsse eine Rolle spielen, gibt es hierzu weitere Untersuchungen. Von [Jab+18] wurden Versuche sowohl tagsüber als auch nachts durchgeführt. Dazu wurde untersucht, ob Brillen und Sonnenbrillen einen Unterschied bei der Bestimmung machen. Aus den Analysen kann abgeleitet werden,

dass die Genauigkeit der Müdigkeitsbestimmung bei Nacht oder mit Brille deutlich abnimmt. Bei den Methoden, welche die Augen durch eine Kamera beobachten, ist die prozentuale Augenschließzeit (englisch: percentage of eye closure - PERCLOS) ein wesentliches Merkmal, was laut [SG10] eine starke Korrelation mit Müdigkeit aufweist. Dieses Merkmal wird auch teilweise als Referenzsignal genutzt, um die Güte der Klassifikation über andere physiologische Messungen auszumachen.
In [LC13] wird es verwendet, um die Daten aus einer Studie einzuordnen. Hierbei wird versucht Signale vom Herzen über eine optische Messung für die Bestimmung der Müdigkeit zu verwenden. Dabei wird untersucht, welche Merkmale aus dem Signal des Herzens zu gewinnen sind. Dazu wird der Einfluss von verschiedenen Merkmalen bei einer Klassifikation der Müdigkeit analysiert. Einige dieser Merkmale wie die Auswertung der Herzraten-Variabilität wird auch in der vorliegenden Arbeit Anwendung finden. Von [Tat+18] wird die Berechnung eines Merkmals auf Basis der Atemfrequenz aus den Rohdaten (R-R Intervallen) beschrieben und mit der direkten Klassifikation der Rohdaten verglichen. Hierbei zeigt sich, das die Berechnung sinnvoll ist und die Klassifikation verbessert. Allerdings handelt es sich hier um eine Vorstudie, in der vier Probanden eine durchschnittliche Gesamtgenauigkeit erreichen, die sich von 55,8 % auf 64,4 % nur leicht verbessert.
In der Arbeit von [Lee+17] werden die Rohdaten mit verschiedenen Sensoren aufgenommen und auf unterschiedliche Arten gefiltert. Dazu wird mittels einer Vorstudie mit vier Probanden evaluiert, ob sich die Signale trotz veränderter Messmethode noch für die Klassifikation von Müdigkeit eignen. In diesem Fall wird ein Verfahren vorgestellt, dass anstatt einer elektrischen Messung des Herzens eine optische Messung vornimmt. Diese ist in der Regel ungenauer und führt zu schlechteren Ergebnissen bei der Klassifikation von Müdigkeit. In diesem Fall sorgen jedoch die vorgestellten Filter für eine gleichwertige oder sogar bessere Klassifikation von Müdigkeit.
In der Studie von [Nau+17] mit 21 Probanden werden verschiedene Merkmale zur Erkennung von Müdigkeit kombiniert. Dazu gehören erstmals auch die Fahrzeit und weitere zusätzliche Informationen wie Kopfbewegungen oder Fahrverhalten. Eine Kombination aus mehreren Merkmalen liefert dabei eine höhere Genauigkeit als einzelne Merkmale. In diesem Studiendesign werden die Probanden jede Minute gebeten, ihre Müdigkeit über Knöpfe am Lenkrad, also durch eine subjektive Messmethode, einzuordnen. Diese Methode, Referenzdaten zu erzeugen, ermöglicht auch, den PERCLOS Wert mit in die Klassifikation einzubeziehen. Dazu wird ein Modell für die Müdigkeit des Fahrers beziehungsweise der Fahrerin erstellt, um eine Aussage darüber zu treffen, wann die nächste Stufe der Müdigkeit erreicht wird.
Eine weitere Möglichkeit, Referenzdaten zur Müdigkeit über eine subjektive Einschätzung zu erlangen, wird von [ÅG90] vorgestellt. Hierbei handelt es sich um eine neunstufige Skala. Bei diesen Untersuchungen werden allerdings nur fünf dieser Stufen mit beschreibendem Text (wie zum Beispiel "weder wach noch schläfrig") versehen. Von [Bau+01] wird eine Beschreibung für jede Stufe der Skala ergänzt. In der von [Ake+14] vorgestellten Zusammenfassung von Versuchen kann gezeigt werden, dass sich diese Skala auch bei Fahrversuchen eignet, wobei zusätzlich zwischen Fahrten im Fahrsimulator sowie Realfahrten unterschieden wird. Dabei erreichen

Fahrten im Fahrsimulator maximal am Ende einen Skalenwert von etwa 8, der höher ist als bei Realfahrten, welche einen Wert von 7 erreichen. Dabei ist der Wert 7 auf der Skala im Englischen mit dem Text "Sleepy - no effort" und der Skalenwert 9 mit dem Text "Fighting sleep, effort keep awake" definiert. Zwischen den beiden Werten ist diese Skala nicht beschriftet. Diese etablierte Skala wird auch in dem Hauptversuch der vorliegenden Arbeit verwendet, da die vorher genannten Veröffentlichungen eine erfolgreiche Bewertung der Müdigkeit zeigen. Sie ist ins Deutsche übersetzt und mit einer Beschriftung an jedem Skalenwert versehen worden. Zu beachten ist, dass die hohen Skalenwerte in der genannten Arbeit nur nachts erreicht werden, wohingegen in der vorliegenden Arbeit der Versuch tagsüber stattfand. Für diesen Zeitraum sind laut der genannten Arbeit eher Werte auf der Skala zwischen 3 und 6 angegeben. Welchen Einfluss diese Selbsteinschätzung auf die Fahrtauglichkeit hat, kann der genannten Arbeit nicht entnommen werden.
Im folgenden Kapitel 2 werden die in der hier vorliegenden Arbeit benötigten Effekte der Psychologie und Physiologie näher erläutert. Dabei wird Bezug auf verschiedene Fahrerzustände, die psychologische Situation beim automatisierten Fahren und die physiologischen Reaktionen der relevanten Organe wie Gehirn, Herz sowie Augen genommen. Kapitel 3 behandelt die Grundlagen rund um künstliche neuronale Netzwerke, besondere Architekturen von Netzwerken und die Fusionstheorie. Anschließend werden in Kapitel 4 der Aufbau der Vorstudie sowie deren psychologische Auswertung des Versuchs dargestellt. Anschließend wird in Kapitel 5 die größere Hauptstudie vorgestellt. Hierbei werden die Themen Design und Durchführung einer Vigilanzstudie und dessen psychologische Auswertung beleuchtet. Im Kapitel 6 werden die Strukturen der Modelle basierend auf verschiedenen Fusionsstufen und deren Training beschrieben. In Kapitel 7 folgt die Auswertung der Modelle auf Basis von Testdatensätzen. Dabei werden zunächst die Bewertungskriterien vorgestellt, sodass im Anschluss jedes Modell für sich im Hinblick auf die Performanz evaluiert werden kann. Dazu wird in diesem Kapitel noch die Auswertung der Ausfallsicherheit aufgezeigt, welche auf einem speziellen Testdatensatz basiert. Zuletzt werden die Modelle verglichen und es wird eine Einschätzung abgegeben, welches Modell sich am besten für zukünftige Untersuchungen eignen würde. Die wesentlichen Erkenntnisse dieser Arbeit werden im Kapitel 8 zusammengefasst und es wird ein Ausblick gegeben. In der vorliegenden Arbeit wird zugunsten der Leserlichkeit an vielen Stellen lediglich die männliche Form verwendet, dies soll jedoch andere Geschlechter und Identifikationen nicht ausschließen.

2

Psychologie und Physiologie bei längerer automatisierter Fahrt

Die Fahrt in einem automatisierten Fahrzeug wird zu neuen psychologischen Situationen und Aufgaben für den Fahrer führen. Diese Veränderung geht mit der steigenden Automatisierung einher, da immer mehr Aufgaben von dem Fahrzeug in vielen Situationen übernommen werden können, aber der Fahrer zunächst weiterhin als Rückfallebene zur Verfügung stehen muss. Die neue beziehungsweise veränderte psychologische Situation wird sich auch in den physiologischen Eigenschaften beim Menschen gemäß der Psychophysiologie widerspiegeln. Nachfolgend wird die psychologische Situation näher beschrieben, die verschiedenen Zustände eines Fahrers erläutert und gezeigt, welche messbaren physiologischen Effekte damit in Beziehung stehen.

Psychologische Situation

Durch [Wee+98] wird deutlich, dass die Aufmerksamkeit ein hypothetisches Konstrukt ist, ebenso wie zum Beispiel die Intelligenz oder Persönlichkeit. Aufmerksamkeit ist [Wee+98] zufolge ein Konstrukt, welches jedem bekannt ist. Ein Konsens über eine einheitliche Definition besteht allerdings noch nicht. Aus schlafmedizinischer Sicht wird das Konstrukt der Aufmerksamkeit in fünf Kategorien differenziert: phasische sowie tonische zentralnervöse Aktivierung, selektive und geteilte Aufmerksamkeit sowie Vigilanz.
Allgemein kann die **zentralnervöse Aktivierung** (im Englischen auch "alertness" oder "arousal level") als eine Art Körperfunktion gesehen werden. Diese ist im Unterbewusstsein verankert und eine Eigenschaft eines Individuums, welche von verschiedenen Faktoren abhängig sein kann. So spiegelt sich das Aktivitätsniveau in verschiedenen physiologischen Merkmalen wider, wie zum Beispiel in der elektrischen Hirnaktivität, der Herzfrequenz, des Hautwiderstandes und der Pupillenweite. Wenn in der vorliegenden Arbeit von Aktiviertheit oder Aktivierung gesprochen wird, ist die zentralnervöse Aktivierung gemeint.
Die **tonische** Komponente der zentralnervösen Aktivität beschreibt die allgemeine Wachheit und Effekte, die durch den circadianen Rhythmus hervorgerufen werden. Der circadianer Rhythmus ist der innere Tagesrhythmus eines Individuums, welcher

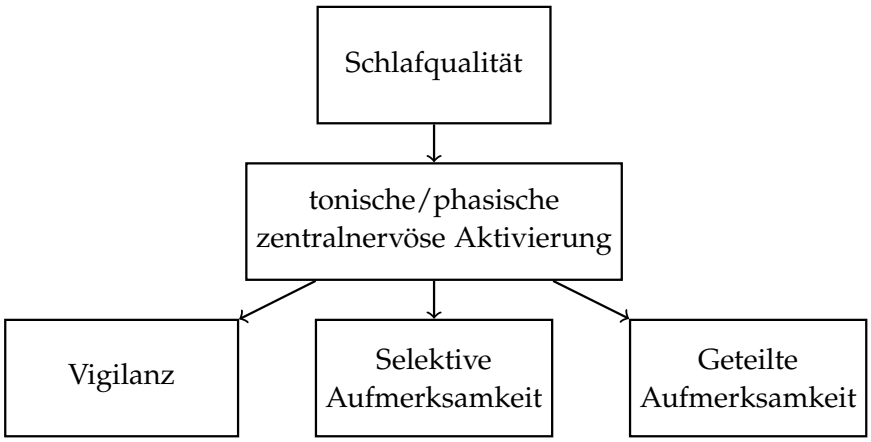


Abbildung 2.1.: Zusammenhang zwischen Schlafqualität und den bewusst gesteuerten Aufmerksamkeitskomponenten aus [Wee+98]

durch die Umwelteffekte von Tag und Nacht geprägt ist. Somit beschreibt er, wie wach sich ein Proband fühlt, abhängig von der Tageszeit und dem vorangegangenen Schlaf. Im Zusammenhang der vorliegenden Arbeit wird versucht, diese Art der Aktivierung durch Fragebögen zu erörtern. Anders als bei schlafmedizinischen Untersuchungen, bei denen bewusst Schlaf entzogen wird, soll hierbei keine Maßnahme zur Veränderung der tonischen zentralnervösen Aktivität vorgenommen werden.

Die **phasische** zentralnervöse Aktivierung ist die Reaktion auf einen Reiz, zum Beispiel ein Zuruf, der es ermöglicht die Aktivierung zu erhöhen. So kann ebenfalls die allgemeine Wachheit also tonische Aktivierung etwas erhöht werden. In der vorliegenden Arbeit wird diese Art der Aktivierung mit Reizen angesprochen, die in einer Fahrsituation implementiert sind. Dabei wird nur mit schwachen und seltenen Reizen gearbeitet, weil dann die Voraussetzung für eine Vigilanzaufgabe gegeben ist.

Die Aufmerksamkeitsaspekte selektive und geteilte Aufmerksamkeit sowie Vigilanz sind der zentralnervösen Aktivierung unterlagert. Dies meint in diesem Fall, dass die zentralnervöse Aktivierung auf all diese Konstrukte einen Einfluss nehmen kann. Zudem stehen diese genannten Aufmerksamkeitsaspekte unter der bewussten Kontrolle der Probanden.

Mit **selektiver Aufmerksamkeit** ist die Fähigkeit gemeint, aus allen auftretenden Reizen diejenigen auszuwählen, die für die Situation notwendig sind. Dieses Konstrukt ist ähnlich dem Filterprinzip, welches von [Haa07] bereits beschrieben wurde. Dabei geht es eher um kurze Situationen, wie zum Beispiel eine Aufgabe der Mathematik, die in einer lauten Umgebung gelöst werden soll. Hierbei ist es wichtig, die irrelevanten Einflüsse heraus zu selektieren und die Aufmerksamkeit für eine gewisse Zeit zu bündeln. In der vorliegenden Arbeit wird versucht, solche Reize so gering wie möglich zu halten und die Aufgabe ist zudem so lang, dass eine Fokussierung auf diese Aufgabe schwer fallen sollte. Diese Komponente der Aufmerksamkeit wird eher bei einer Übernahmesituation eine größere Rolle spielen, welche im Folgenden nicht weiter untersucht wird.

Im Gegensatz zur selektiven Aufmerksamkeit handelt es sich bei der **geteilten Aufmerksamkeit** um die Fähigkeit, mehrere Reize sowohl seriell als auch parallel zu verarbeiten. So kann das manuelle Fahren als Beispiel für diese Art von Aufmerksamkeit gesehen werden. Hierbei muss die Umfelderfassung parallel zu den motorischen Bewegungen zur Fahrzeugführung ausgeführt werden. Durch das Training in der Fahrschule können viele Komponenten, wie zum Beispiel die Gangwahl, schnell und unbewusst ausgeführt werden. In der vorliegenden Arbeit soll durch eine spezielle monotone Aufgabe, welche in wenigen Reizen resultiert, möglichst wenig der geteilten Aufmerksamkeit angesprochen werden.
Wie bei den ersten beiden unterlagerten Aufmerksamkeitsaspekten deutlich wird, ist die Unterscheidung von der Art der Aufgabe abhängig. Bei der Definition der **Vigilanz** zu Beginn der Arbeit wurde deutlich, dass es hier um eine längere Daueraufmerksamkeitsaufgabe mit schwachen und seltenen Reizen geht. Dabei wird eine erhöhte Reaktionsbereitschaft gefordert, welche zum Beispiel über Reaktionszeiten gemessen werden kann. Hinsichtlich der zentralnervösen Aktivierung, sollte in diesem Fall also weniger die phasische zentralnervöse Aktivierung angesprochen werden. Es steht eher die Präsenz von einer erhöhten tonischen zentralnervösen Aktivität im Fokus.
Das Gegenteil von Vigilanz wird **Hypovigilanz** genannt und beschreibt eine verminderte Reaktionsbereitschaft. Diese kann ebenfalls durch Reaktionszeiten oder ausbleibenden Reaktionen messbar gemacht werden. Eine genauere Beschreibung der Hypovigilanz wird von [Haa07] erarbeitet. In der vorliegenden Arbeit wird der psychologische Zustand der Probanden entweder als vigilant oder hypovigilant eingeordnet, da durch den Gegensatz immer einer der beiden Zustände zutrifft.
Dieser Effekt wurde zunächst in der Arbeitswissenschaft systematisch untersucht. Hierbei standen Tätigkeiten, wie das Beobachten von Radarschirmen im maritimen Bereich oder Flugwesen, im Fokus. So wird von [Mac48] gezeigt, dass nachlassende Reaktionsfähigkeit auftreten kann, wenn die Signalentdeckungsaufgaben mit geringen Reizen einhergehen. Bei dieser Untersuchung besteht die Aufgabe darin, einem Lichtpunkt auf einer Uhr mit zwölf Positionen zu folgen. Wenn dieser Lichtpunkt eine Position überspringt, soll der Proband dies mit einem Tastendruck angeben. Zusätzlich zu der Messung von Reaktionszeiten können auch andere Merkmale des Probanden gemessen werden. [Zha+17a] untersuchte die Auswirkung auf die Pupillenweite bei einer ähnlichen optischen Vigilanzaufgabe und nutzte außerdem subjektive Bewertungen der Probanden.
In der Arbeitswissenschaft wird die Vigilanz später auch in anderen Bereichen untersucht. So liegt der Schwerpunkt der Untersuchung von [Haa07] auf der Automatisierung im Flugverkehrswesen. Hierbei besteht die Schwierigkeit darin, dass Piloten eine längere Zeit den Autopiloten durch die Begutachtung von Instrumenten überwachen müssen. Bei dieser Arbeit wird eine adaptive Automatisierung vorgeschlagen, welche Aufgaben zuweist oder übernimmt, je nachdem wie der Aktivierungszustand des Piloten ist. Hierfür werden, wie in der vorliegenden Arbeit, psychophysiologische Merkmale erfasst und dann für eine Adaption weiterverwendet.
Die Aufgabe, um die es in der vorliegenden Arbeit primär geht, ist das Führen eines Kraftfahrzeugs. Aus psychologischer Sicht wird diese Aufgabe zum Beispiel

von [Jür+18] untersucht. Hierbei geht es besonders um die Unterschiede zwischen erfahrenen und unerfahrenen Kraftfahrzeugführern. Am Rande wird in der zitierten Arbeit aber auch eine grobe Einschätzung über zeitliche Abläufe gegeben, die beim manuellen Fahren auftreten. Gemeint ist hiermit, in welchen Abständen ein Mensch auf die typischen Reize beim Fahren reagiert und daraufhin agiert. Es wird deutlich, dass sich diese Vorgänge in einer Zeitspanne von 0,5 s bis 1 s abspielen. Typischerweise vergeht zwischen zwei Korrekturen bei der Querführung, also dem Spurhalten, etwa 0,7 s, wobei sich dieser Wert vermutlich nicht auf die in der vorliegenden Arbeit untersuchte Autobahnfahrt übertragen lässt. Bei einer Geschwindigkeit von 130 $\frac{\text{km}}{\text{h}}$ bewegt sich ein Auto etwa 25 m in Längsrichtung fort. Fährt das Fahrzeug nicht exakt parallel zur Spurmarkierung, sondern mit einem Winkel von etwa 4 Grad auf diese zu, so wird in der entsprechenden Zeitspanne die Fahrspurmarkierung bei einer 3 m breiten Spur bereits erreicht.

Der Fahrer muss also in kurzen Abständen auf den Reiz reagieren, der durch das Verlassen des Fahrzeugs aus der Fahrspurmitte resultiert. Ebenso wird in etwas längeren Zeitintervallen die Geschwindigkeit des Fahrzeugs kontrolliert und angepasst. Dazu muss zu jeder Zeit der Verkehr sowohl vor dem Fahrzeug als auch durch die Spiegel nach hinten kontrolliert werden.

Das manuelle Fahren fällt somit eher unter die Daueraufmerksamkeitsaufgaben mit einer hohen **Reizrate**. Die Reizrate beschreibt eine Zeitspanne zwischen den Reizen, auf die ein Fahrer reagieren sollte. Bei den frühen Untersuchungen von [Mac48] liegt diese Zeitspanne zwischen 45 s und 5 Minuten, wobei normale Signale, auf die nicht reagiert werden soll, jede Sekunde auftreten. Bei einer Automatisierung ab SAE Level 2 können beide Aufgaben zur Fahrzeugführung entfallen. Es kann dazu führen, dass die Reizrate für den Menschen wesentlich geringer ausfällt, da seltener auf ein Ereignis reagiert werden muss. Dies gilt bei geringem Verkehrsaufkommen oder einer Fahrt auf der Autobahn, bei der es keine Kreuzungen bzw. keinen Gegenverkehr gibt. In der vorliegenden Arbeit wird die Reizrate nicht direkt untersucht, jedoch unterschiedlich groß gewählt. Bei der Machbarkeitstudie wird etwa alle zwei bis drei Minuten ein Reiz präsentiert. Die gewöhnlichen Signale, auf die nicht reagiert werden sollen, erfolgen jedoch in einem Abstand von etwa 25 s. In der folgenden Vigilanzstudie variiert die Reizrate stark, denn es gibt eine Zeitspanne von 20 Minuten ohne Reiz. In den meisten Fällen liegen die Reize jedoch drei bis sechs Minuten auseinander.

In der Literatur sind schon zahlreiche Untersuchungen zur Vigilanz im Zusammenhang mit dem Führen eines Fahrzeugs zu finden. [Cab+16] gibt einen Überblick über eine Vielzahl an Veröffentlichungen und gruppiert diese ein. Bei der Integration der Vigilanzaufgabe können zwei Eigenschaften unterschieden werden, die sich auf die Fahraufgabe beziehen. Zum einen können die Reize der Vigilanzaufgabe in die Fahraufgabe integriert werden und zum anderen kann die Reaktion in Bezug zur Fahraufgabe erfolgen. Die Integration des Reizes wird zum Beispiel durch seltenes Verlassen der Spur aufgrund von simuliertem Seitenwind oder durch Überwachung der Längsführung erreicht. Ein Beispiel hierfür findet sich in [Chu+14], bei dem das Verlassen der Spur umgesetzt wird. Der Fokus dieser Untersuchung liegt jedoch darin, ob eine kinetische Rückmeldung vom Fahrsimulator einen Unterschied in der Gehirnaktivität des Probanden erzeugt und wo diese auftritt.

Hierbei wurde festgestellt, dass eine kinetische Rückmeldung vom Fahrsimulator die Reaktionsfähigkeiten der Probanden verbessern kann. Dies ist ein wichtiger Hinweis, denn in der vorliegenden Arbeit wird die Untersuchung in einem statischen Fahrsimulator durchgeführt. Wird der Reiz nicht in die Fahraufgabe integriert, wird oftmals eine auditive Vigilanzaufgabe verwendet. So wird von [Kör+15] eine Vigilanzaufgabe in Form von verschiedenen Tönen umgesetzt. Hierbei ist das Reizsignal ein Ton mit einer höheren Frequenz als bei den im Abstand von sechs Sekunden abgespielten Signaltönen, auf welche nicht reagiert werden sollte. Auch hier werden subjektive Fragebögen zusätzlich zur Bewertung herangezogen.

In dieser Arbeit konnte festgestellt werden, dass die Aufgabe zu leicht ausgefallen ist, denn die Versuchszeit hat keinen signifikanten Einfluss auf Reaktionszeiten und ausbleibende Reaktionen gezeigt. Lediglich in der Untersuchung der Blinzelfrequenz konnte über die Versuchszeit ein signifikanter Anstieg festgestellt werden. Neben der Integration der Reize in die Fahraufgabe, ist es auch möglich, die Reaktionsmöglichkeiten mit der Fahraufgabe zu kombinieren. Exemplarisch dafür steht die Reaktion mit Bedienelementen im Fahrzeug, wie in der Arbeit von [Rog+02] den Scheibenwischer-Hebel zu betätigen oder das Verlassen der Spur mit Lenkbewegungen zu korrigieren, wie es in der schon genannten Arbeit von [Chu+14] beschrieben ist. Neben der Unterscheidung, wie sehr die Vigilanzaufgabe in die Fahraufgabe integriert ist, kann auch noch zwischen Real- und Simulator-Fahrten unterschieden werden. Es konnte nachgewiesen werden, dass etwa 70 % der Untersuchungen in einem Simulator stattgefunden haben. Eine Untersuchung der Vigilanz im Realverkehr in einem automatisierten Fahrzeug, welches sowohl Längs- als auch Querführung übernimmt, wird beispielsweise von [Sta+19] vorgestellt. Hierbei wurde eine auditive Vigilanzaufgabe verwendet, die, wie oben beschrieben, nicht in die Fahraufgabe integriert worden ist. Mit dieser Untersuchung wird die Vigilanz unter verschiedenen Bedingungen betrachtet. In diesem Fall variiert die Automationsstufe, die Komplexität der Fahrsituation sowie die Erfahrung mit der Automatisierung. So wird der Einfluss von unterschiedlich schwierigen Situationen sowie den verschiedenen Erfahrungswerten auf die kognitive Belastung untersucht. Dabei geht es weniger als in der vorliegenden Arbeit um die Erfassung der kognitiven Belastung während der Fahrt mit Sensoren, da diese ausschließlich über Fragebögen und einen Reaktionstest erfasst wird. Dabei konnte festgestellt werden, dass bei unerfahrenen Fahrern auch eine höhere kognitive Belastung zu messen ist. Eine andere Möglichkeit, Bedingungen nahe an Realfahrten in einem automatisierten Fahrzeug zu erzeugen, ist die Methode des Wizard-of-Oz-Fahrzeugs. Hierbei wird eine automatisierte Fahrt in einem Realfahrzeug dadurch simuliert, dass eine zweite Person das Fahrzeug führt, die nicht für den Probanden sichtbar ist. So wirkt es für den Probanden, als würde das Fahrzeug automatisiert fahren. Dabei wird dem Probanden erklärt, dass das Fahrzeug durch eine Automation gefahren wird. Eine Vigilanzuntersuchung in einem solchen Fahrzeug wird von [Fel+18b] vorgestellt. Hierbei werden für die Ermittlung der Vigilanz ein Eye-Tracking-System, Herzraten, Fragebögen und weitere Merkmale verwendet. In dem Vorversuch von [Fel+18b] geht es, ähnlich zu dem Vorversuch in der vorliegenden Arbeit, um die Erkennung von bestimmten Schildern auf der Autobahn. Dabei sollte, anders als in der hier

vorliegenden Arbeit, nicht direkt auf solche Reize reagiert werden. Die Aufgabe bestand darin, nach der Fahrt ungewöhnliche Situationen zu beschreiben und aus einer Liste die Elemente auszuwählen, die aufgetreten sind. Hierbei wurde festgestellt, dass die Probanden diese Reize größtenteils nicht wiedergeben konnten. Am Ende der automatisierten Fahrt gab es ein Hindernis, auf das reagiert werden sollte. Dabei wurde die Fähigkeit getestet, nach langer Automatisierung, wieder manuell fahren zu können. Es konnten jedoch keine wesentlichen Erkenntnisse in Bezug zur Vigilanz aus dem Vorversuch gezogen werden. Im Hauptversuch wird der Vergleich zwischen einer Hypovigilanz-Gruppe und einer Gruppe mit fahrfremden Tätigkeiten unterschieden. In diesem wurden ebenfalls, wie in der hier vorliegenden Arbeit, Daten von Augenschließungen für die Bestimmung von Müdigkeit und Vigilanz herangezogen. Es wurde festgestellt, dass es nicht möglich ist, Vigilanz und Müdigkeit allein durch statistische Untersuchungen von den Augenaufnahmen voneinander zu trennen. Der Unterschied zu der vorliegenden Arbeit ist, dass in der zitierten Arbeit keine automatisierte Auswertung der aufgenommen Daten vorgenommen wird. Somit kann zwar eine statistische Auswertung erfolgen, jedoch ist diese nicht in einem Modell beschrieben, welches den Zustand des Fahrers selbständig angeben kann.

Fahrerzustände

Der Fahrerzustand wird von [Mar+18] in drei Kategorien unterteilt: den sensorischen, motorischen sowie kognitiven Zustand. Insgesamt wird hierbei versucht, Zusammenhänge zu dem Aktionspotential herzustellen und diese dann in Kategorien aufzuteilen, um sie besser zu verstehen.
Der **sensorische** Zustand bezieht sich dabei auf die Fähigkeit, durch Wahrnehmung eine Situation aufzunehmen. Bezogen auf den Fahrer bedeutet dies, ob dieser in der Lage ist, etwas zu hören, zu sehen oder mit anderen sensorischen Eigenschaften etwas aufzunehmen. Dieser Zustand könnte zum Beispiel durch eine Eye-Tracking-Brille aufgenommen und analysiert werden. Diese Untersuchung steht in der vorliegenden Arbeit jedoch nicht im Fokus.
Der **motorische** Zustand beschreibt die Möglichkeit des Fahrers, durch Aktionen situationsbezogen einzugreifen. Dies ist maßgeblich von der Körperhaltung abhängig. Es wird in diesem Zustand beispielsweise unterschieden, ob die Hände am Lenkrad sind oder nicht, da dies ebenso Einfluss auf das Aktionspotential hat. Die Pose des Fahrers kann auf verschiedene Weise aufgenommen werden. Dieser Zustand spielt in der vorliegenden Arbeit eine eher kleinere Rolle, denn durch Sitzdrucksensoren wird dieser Zustand aufgenommen und fließt auch mit in die Berechnung des Fahrerzustands ein, jedoch werden diese Daten nicht tiefer analysiert.
Der **kognitiven** Zustand deckt die Merkmale ab, die innerhalb des Gehirns des Fahrers vorgehen und so auf das Aktionspotential einwirken. In der vorliegenden Arbeit wird die Gehirnaktivität aufgezeichnet, welche einen Aufschluss über diesen Zustand erbringen soll. Da dieses Signal schwierig zu analysieren ist, wird versucht mit Fragebögen Aufschluss über diesen Zustand zu gewinnen. Ebenso kann sich die Aktivität im Gehirn noch in anderen physiologischen Signalen widerspiegeln, sodass

in der vorliegenden Arbeit weitere Merkmale aus Aufnahmen des Herzens und der Augen berechnet werden.

Alle Untergruppen des Fahrerzustands sind miteinander verbunden. So hat zum Beispiel die Wahrnehmung einen Einfluss auf den kognitiven Zustand, da sensorisch aufgenommene Signale zu einer Verarbeitung im Gehirn führen. So kann der kognitive Zustand auch Einfluss auf den motorischen Zustand haben, indem durch die kognitive Anstrengung eine andere Pose eingenommen wird oder bestimmte Bewegungen nicht direkt umsetzbar sind. In der oben genannten Arbeit werden die Aktiviertheit (Arousal Level) und Motivation als einzelner Block dargestellt. Alle Untergruppen sind von der Aktiviertheit und der Motivation abhängig, sodass eine Veränderung der Zustände von diesen beiden Nebenbedingungen abhängt. Neben diesen drei Untergruppen wird in anderen Arbeiten auch noch in weitere Zustände unterschieden. So ist zum Beispiel auch von dem **emotionalen** Zustand die Rede, welcher in den oben genannten Klassen als Untergruppe des kognitiven Zustands verstanden werden kann. Dieser Zustand wird in der vorliegenden Arbeit nicht im Detail beachtet.

Physiologische Eigenschaften

Die Psychophysiologie beschreibt, wie der sensorische, motorische und kognitive Zustand im Zusammenhang mit der Hirntätigkeit, dem Kreislauf, der Atmung und der Hormonausschüttung steht. Dabei existieren jedoch starke individuelle Unterschiede, sodass oftmals der Zusammenhang lediglich qualitativ bestimmt werden kann. So gibt es zu verschiedenen Körperfunktionen bereits Thesen für den Zusammenhang zum psychologischen Zustand der Vigilanz.

Bei der Messung dieser Körperfunktionen kann zwischen invasiven und nicht invasiven Methoden unterschieden werden. Hierbei ist nicht wie in der Medizin gemeint, dass eine Messmethode in den Körper des Probanden eindringt, sondern dass der Sensor in eine normale Fahrsituation in einem Fahrzeug eindringen beziehungsweise stören würde. Beispielsweise wäre das Anlegen eines Sensors am Kopf oder am Arm in diesem Fall eine eher invasive Messmethode, die Beobachtung durch eine Kamera hinter dem Lenkrad eine eher nicht invasive Messmethode. [Leo+18] fasst nicht invasive Messmethoden im Kontext des Fahrzeugführens detailliert zusammen. Dabei wird auch Bezug auf Sensoren im Lenkrad genommen, die in der vorliegenden Arbeit nicht von Interesse sind, da bei einer Fahrt mit einem hohen Automatisierungsgrad (ab SAE Level 2) das Lenkrad nicht zwangsläufig berührt werden muss. Fast alle Messmethoden werden als Konzept dargestellt und in den meisten Fällen noch nicht als fertiges Messsystem angeboten. In der vorliegenden Arbeit soll es nicht um Sensorentwicklung gehen, sodass auf Methoden zurückgegriffen wird, die vollständig entwickelt verfügbar sind. Somit mussten einige der benötigten Körperfunktionen mit invasiven Messmethoden aufgenommen werden, welche in der Zukunft durch nicht invasive Methoden abgelöst werden könnten. Es wurde für die Aufnahme des Herzens zum Beispiel ein Brustgurt mit Kontakt zur Haut verwendet, anstatt diese Information aus der Ferne aufzunehmen.

Gehirn

Von der Gehirnaktivität ausgehend lassen sich Aktivitäten in der Hirnrinde (Kortex) durch an der Schädeloberfläche auftretende Spannungsunterschiede (Gehirnaktivität) bestimmen (EEG – Elektroenzephalografie). Ursprünglich wird in der Medizin das EEG mit Nadeln auf einer fortlaufenden Papierrolle aufgezeichnet. Experten werten dann die Spannungssignale, wie zum Beispiel von [Zsc02] gezeigt, im Zeitbereich aus. Ohne eine genaue Spektralanalyse durchzuführen, ist es möglich, verschiedene Aktivitäten anhand unterschiedlicher Muster im Zeitsignal zu unterscheiden. Damit können auch direkt Störungen und krankhafte Signale eingeordnet werden. Um nun Signale für eine objektive Messung des psychologischen Zustands eines Menschen zu erhalten, wird eine Spektralanalyse durchgeführt, wie es [Nov19] für verschiedene Anwendungen zeigt. Hierbei wird das Frequenzspektrum der Spannungswerte in fünf überlappende Frequenzbänder [Int15] aufgeteilt: Delta (δ) [1-4 Hz], Theta (θ) [4-8 Hz], Alpha (α) [7,5-13 Hz], Beta (β) [13-30 Hz] und Gamma (γ) [30-44 Hz].

Die einzelnen Frequenzen lassen sich, wie von [Zsc02] gezeigt, mit verschiedenen Charakteristika in Zusammenhang bringen. So wird bei dem hier vorliegenden Versuch erwartet, dass eine höhere θ-Aktivität zu finden ist, weil diese in Phasen von Vigilanzminderung auftreten kann. Ebenso ist es möglich, dass vorübergehend eine erhöhte β-Aktivität festzustellen ist, weil diese mit psychischer Belastung und verminderter Vigilanz (subvigile β-Aktivität) in Verbindung gebracht werden kann. Eine erhöhte γ-Aktivität soll dabei eher bei einem Wachzustand vorgefunden werden. Es gibt Untersuchungen, die den Zusammenhang zwischen Vigilanz und der Gehirnaktivität erforschen. So gibt [Arm+16] einen Ansatz zur Klassifikation zwischen zwei Zuständen: zum einen ein entspannter Zustand mit offenen Augen und zum anderen das Ausführen einer Vigilanzaufgabe. Hierbei wurden verschiedene Merkmale herangezogen, darunter auch die zuvor erwähnten Frequenzbänder und Kombinationen aus diesen. Dann wird mit einer Auswahl an Merkmalen eine Support Vector Maschine trainiert, welche eine Genauigkeit von etwa 95 % erzielt. Dabei handelt es sich um das Ergebnis der binären Klassifikation (zwischen vigilant und hypovigilant) im Mittel über fünf Probanden, wobei andere Teile der Datenreihe von den gleichen Probanden (wie beim Training) verwendet werden. Im Vergleich zu der hier vorliegenden Arbeit, ist die Situation in der zitierten Arbeit sehr einfach gehalten und es gibt keinen Bezug zum Führen eines Fahrzeugs oder zur Überwachung einer Automation. Von [Haa07] wird die Messung der Gehirnaktivität verwendet, um eine Adaption der Automatisierung eines Autopiloten im Flugwesen umzusetzen. Hierbei wird der von [Pop+95] vorgestellte EEG-Engagement-Index (EEI $= \frac{\beta}{\alpha+\theta}$) verwendet, um den Grad der Beanspruchung festzustellen. Die Berechnungsgrundlage der notwendigen Frequenzbandsignale wird in dieser Arbeit nicht genauer beschrieben. Es wird lediglich festgehalten, dass die EEG-Signale mit einer Abtastung von 200 Hz digitalisiert werden. Die Berechnung der Frequenzbänder bedingt eine Fouriertransformation, welche ein gewisses Intervall an Daten berücksichtigt. Es wird erwartet, dass der EEI geringe Werte für eine geringe Belastung hervorbringt. In den Ergebnissen wird bestätigt, dass sich dieser Index für die Bestimmung der psychologischen Belastung gut eignet. Hier ist ebenfalls kein Bezug zum Führen eines

Fahrzeugs vorgestellt, denn beide gerade zitierten Arbeiten haben den Hintergrund in der Automatisierung im Flugwesen. Eine weitere Möglichkeit zur Nutzung der EEG zur Bestimmung der Vigilanz stellt [Xia+17] vor. Um die EEG-Daten zu verarbeiten, wird in diesem Fall eine Wavelet-Transformation herangezogen, um die Frequenzbänder zu berechnen. Dann wird ebenfalls eine Support-Vector-Maschine mit den gleichen drei Frequenzbändern (α, β, θ) einzeln als Merkmale für die Klassifikation verwendet. Mit dieser Methode kann eine Genauigkeit von etwa 90 % erreicht werden, wobei nicht angegeben wird, ob die unterschiedlichen Zustände gleichmäßig in den Testdaten vertreten waren. Zudem wird bei der Klassifikation nicht, wie in der vorliegenden Arbeit, angenommen, dass der psychologische Zustand auch zeitabhängig ist.
Eine andere Möglichkeit, die Signale des Gehirns zu beobachten, ist die Nahinfrarotspektroskopie (NIRS). Diese Methode basiert nicht auf elektrischen Effekten, sondern verwendet ein optisches System, mit dem die Durchblutung auch in tieferliegenden Schichten des Gewebes untersucht werden kann. Hiermit ist es möglich, eine wesentlich detailliertere räumliche Auflösung der Gehirnaktivität zu erhalten. Von [Sib+17] wird eine solche Messmethode bei der Untersuchung von der Gehirnaktivität bei Fahrten in einem Fahrzeug mit verschiedenen Automatisierungsstufen verwendet. Die Beeinträchtigung durch die Messapparatur, welche in der zitierten Arbeit nur mit Kabeln und einer Kopfhaube funktioniert, ist ein Nachteil dieser Methode. Dadurch wird die Messung wesentlich aufwändiger, als es mit der EEG möglich ist. Dazu können Kabel die Aufmerksamkeit der Probanden beeinträchtigen. Aus diesem Grund wird in der vorliegenden Arbeit ein kabelloses System verwendet, um die Gehirnaktivität der Probanden aufzunehmen.

Herz

Das Herz kann bei verschiedenen psychologischen Situationen seinen Rhythmus verändern und anpassen. Dabei sind zwei Maße ausschlaggebend für die Analyse des Herzrhythmus, die Herzfrequenz (HF) und die Herzraten-Variabilität H. Bei der Herzfrequenz (HF) geht es grundlegend um die Häufigkeit von Herzschlägen pro Zeitspanne. In den meisten Fällen bezieht sich dieses Maß auf Herzschläge pro Minute mit der Einheit beats per minute (bpm). Die Herzraten-Variabilität wird, wie in [SG17] zusammengefasst, zum Beispiel durch eine Analyse der Zeitdifferenzen zwischen einzelnen Herzschlägen charakterisiert. Dabei berechnen sich die Zeiten zwischen einzelnen Herzschlägen r_i über eine Auswertung der Kurve des Elektrokardiogramms. Hierzu wird die Zeit zwischen den sogenannten R-Wellen in dem Signal bestimmt (siehe [Gil+19]). So kann der Wert der Herzraten-Variabilität H über eine Zeitspanne Z_H mit den Zeiten zwischen den Herzschlägen r_i wie folgt berechnet werden:

$$H = \sqrt{\sum_{i=0}^{Z_\mathrm{H}} (r_{i+1} - r_i)^2} \tag{2.0.1}$$

Die Zeitspanne Z_H wird je nach Untersuchung unterschiedlich gewählt. Für den medizinischen Bereich (um Krankheiten zu diagnostizieren) wird die H oftmals über

einen Zeitraum von 24 Stunden oder mehr untersucht und eine Zeitspanne Z_{H} von mindestens fünf Minuten gewählt. Daneben gibt es auch schon Untersuchungen zu kürzeren Zeitspannen, die im Englischen "Ultra-Short-Term" genannt werden. Eine Untersuchung hierzu wird von [EF14] vorgestellt, bei der es um Unterschiede bei der Berechnungen von H von Athleten vor und nach einer körperlicher Anstrengung geht. Die Berechnung von H wird durch die Wahl von verschiedenen Zeitspannen Z_{H} variiert, welche zunächst den Standardwert von fünf Minuten einnimmt. Danach werden bei den Berechnungen Zeitspannen von 10, 30 und 60 Sekunden verwendet. So kann festgestellt werden, dass bei dieser Untersuchung eine Zeitspanne Z_{H} von 60 Sekunden sehr gut für eine solche Untersuchung geeignet ist. In der vorliegenden Arbeit wird für die Zeitspanne ebenfalls $Z_{\mathrm{H}} = 60\,\mathrm{s}$ verwendet.Dabei werden die Zeiten zwischen den Herzschlägen (r_i) jede Sekunde von einem Sensor bereitgestellt und somit kann der Wert H auch für jede Sekunde mit dem entsprechenden Intervall neu berechnet werden.

In [KS67] wird gezeigt, dass bei zunehmender mentaler Belastung bei zwei parallelen Aufmerksamkeitsaufgaben (sowohl visuell als auch akustisch) die Herzraten-Variabilität abnimmt. So wird bei diesem Merkmal erwartet, dass der Wert H in der vorliegenden Arbeit im Verlauf des vorgestellten Versuchs zunehmen sollte, da die mentale Belastung als niedrig eingeschätzt wird. In der Untersuchung von [Haa07] wird die Herzfrequenz bei der Überwachung von Autopiloten im Flugzeug beobachtet. Hier wird eine niedrigere Herzfrequenz für die Phasen der Überwachung des Systems bei hoher Automatisierung, im Vergleich zu Phasen mit weniger Automatisierung, festgestellt. Dabei hat weder das Geschlecht noch die Tageszeit einen Einfluss auf diese Unterschiede. Ebenso wird die Herzraten-Variabilität untersucht. Es konnte jedoch kein signifikanter Effekt diagnostiziert werden. Die Tendenz der Herzraten-Variabilität nimmt in mental anstrengenden Phasen des Versuchs allerdings einen geringeren Wert an als in Ruhephasen und bei der Überwachung des Systems. Von [Fel+18b] wird die Herzfrequenz bei einem Vigilanzversuch in einem Fahrsimulator aufgenommen. Bei dem Vorversuch in der genannten Arbeit ist die Datenlage statistisch nicht auszuwerten, da die Anzahl der Probanden zu gering war. Es wird erwartet, dass bei dieser Untersuchung die Herzfrequenz durch die Unterforderung absinkt. Eine leichte Tendenz einer sinkenden Herzrate konnte dabei nach etwa 30 Minuten festgestellt werden. Die Werte stiegen jedoch wieder gegen Ende des Versuchs zur Minute 50. Im Hauptversuch der genannten Arbeit wird dann eine statistische Auswertung der Herzfrequenz über die Zeit beschrieben. Hierbei kann kein signifikanter Unterschied in der Herzfrequenz zwischen den Gruppen (Hypovigilant und fahrfremde Tätigkeiten) festgestellt werden. Dazu wird dann noch die relative Änderung der Herzfrequenz untersucht. In dieser unterscheiden sich die Auswertungen der verschiedenen Gruppen, wobei die Hypovigilanz-Gruppe eine Abnahme der Herzfrequenz zur Mitte des Versuches bei Minute 35 zeigte. Auch hier sind die Ergebnisse nicht signifikant und es wird keine Methode bzw. kein Wert für die Signifikanz angegeben. Anders als in der vorliegenden Arbeit wird in diesem Fall die Herzfrequenz nicht kontinuierlich herangezogen, sondern alle zehn Minuten der Mittelwert über 60 s gebildet und anschließend werden die Mittelwerte verglichen. Dieses Vorgehen könnte dazu führen,

dass schnellere Änderungen der Herzfrequenz nicht berücksichtigt werden können. In der Untersuchung von [Sta+19] werden die Herzfrequenz und Herzraten-Variabilität bei Realfahrten erfasst. In der Auswertung konnte kein signifikanter Effekt gezeigt werden. Auch hier wurde die Herzfrequenz über gewisse Intervalle berechnet. Ebenso brachte eine Untersuchung zu kleineren Intervallen keinen signifikanten Effekt mit sich, wobei nicht genauer beschrieben wurde, welche Zeitintervalle in diesem Fall untersucht wurden.

Augen

Die Augen spiegeln durch verschiedene Signale eine Änderung des psychologischen Zustands wider. Sowohl die Augenbewegung und Fixation als auch die Pupillenweite und Schließdauer sind hilfreich für die Bestimmung eines Zusammenhangs zwischen einer physiologischen Veränderung und der kognitiven Belastung des Fahrers [Mar+15]. Die **Fixation** beschreibt ein längeres Verharren der Blickrichtung auf einen bestimmten Punkt. Die normalen Augenbewegungen laufen in einem Zeitrahmen von 100 bis 200 Millisekunden ab. Bei Fixationen verharrt das Auge über 250 bis 300 Millisekunden. In der vorliegenden Arbeit wird darauf verzichtet, eine so genaue Beobachtung der Augen durchzuführen, weil diese Auswertung sehr aufwändig ist und zusätzlich zu ablenkenden Messmethoden führen könnte (wie das Tragen einer Eye-Tracking-Brille). Dazu sind solche Methoden in Serienfahrzeugen schwieriger umzusetzen und damit teuer. In der vorliegenden Arbeit liegt der Fokus auf der Augenschließdauer, da sie leicht aus einiger Entfernung berührungslos durch eine Kamera bestimmbar ist. Üblicherweise wird, wie in [WE94] gezeigt, die prozentuale Zeit, in der das Auge zu mindestens 80 % geschlossen ist (englisch: PERcentage of Eyelid CLOSsure – PERCLOS), als Maß herangezogen. Dabei wird als Zeitintervall 60 s vorgeschlagen und zudem eine Korrelation zwischen dem Ansteigen des Wertes und der Müdigkeit festgestellt. Dieser Zusammenhang wird in [SG10] konkret bei dem Führen eines Fahrzeugs untersucht und mit anderen Signalen, unter anderem der Gehirnaktivität, kombiniert. Dabei kann ermittelt werden, dass die prozentuale Augenschließdauer (PERCLOS-Wert) mit der Müdigkeit beim Fahren eines Fahrzeugs korreliert. Dazu wird die Müdigkeit, neben dem PERCLOS-Wert, über eine subjektive Einschätzung des Probanden durch die Karolinska Sleepiness Scale (KSS) [ÅG90] bestimmt, sodass eine Korrelation zwischen den verschiedenen Messmethoden gezeigt werden kann. Die Skala reicht von 1 bis 9, wobei 1 mit *Extrem wach* und 9 mit *Sehr schläfrig, kann nur mit großer Mühe wach bleiben; kämpfe gegen den Schlaf* beschriftet ist. Dazwischen sind drei weitere Werte mit einer Beschriftung versehen: 3 ist mit *Wach*, 5 ist mit itWeder wach noch schläfrig und 7 mit *Schläfrig, aber kann noch ohne Mühe wach bleiben* versehen. In der zitierten Arbeit ist die Skala mit englischen Begriffen versehen, die im Rahmen der vorliegenden Arbeit ergänzt und übersetzt wurden. Im Vergleich zu der Gehirnaktivität ist die Korrelation zwischen Messwerten und Müdigkeit bei dem PERCLOS-Wert jedoch schwächer ausgeprägt. Eine Auswertung des PERCLOS-Wert im Zusammenhang mit dem automatisierten Fahren wird von [Fel+18b] vorgestellt. Die Auswertung des PERCLOS-Wertes über die Versuchsgruppe ergab einen Anstieg des Mittelwertes zur Hälfte der

Versuchszeit bei 35 Minuten mit einer höheren Varianz und danach einen Abfall auf das Anfangsniveau. Eine Auswertung einzelner Probanden hat gezeigt, dass sich der Wert nur für etwa die Hälfte der Gruppe veränderte. Dabei gab es bei denjenigen, bei denen ein Unterschied existiert, verschieden starke Ausprägungen der Änderungen. Diese traten jedoch eher in den ersten beiden Dritteln des 60-minütigen Tests auf. Durch die Varianzanalyse konnten jedoch keine signifikanten Unterschiede mit der Versuchszeit oder den Altersgruppen der Probanden ausgemacht werden. Eine weitere Art der Augenauswertung ist die Blinzelhäufigkeit, die hier ebenfalls untersucht wurde. Hierbei wurden ebenfalls keine signifikanten Unterschiede mit der Versuchszeit deutlich. Darüber hinaus wurden noch überlange Lidschlussvorgänge betrachtet. Im Gegensatz zu normalen Augenschließungen von durchschnittlich 300 bis 400 Millisekunden werden diese durch eine Schlusszeit von mehr als 500 Millisekunden charakterisiert. Hierbei zeigt sich ein Anstieg dieses Parameters über die Versuchszeit, der bis Minute 35 anhält, bei Minute 45 abfällt und dann noch einmal etwas ansteigt. Sowohl zur Signifikanz der Auswertung als auch zur Varianz werden in diesem Fall keine Angaben gemacht. Die Blinzelfrequenz, die Lidschlussdauer, der PERCLOS-Wert und der Pupillendurchmesser wurden von [Kör+15] bei einer Vigilanzuntersuchung herangezogen. Hierbei konnte bei der Lidschlussdauer ein signifikanter Anstieg von etwa 172 ms auf etwa 184 ms im Verlauf der Versuchszeit festgestellt werden. Bei der Blinzelfrequenz wurde ebenfalls ein signifikanter Anstieg von etwa 0,3 auf 0,44 Lidschlüssen pro Sekunde in der Versuchszeit ermittelt. Ebenso konnte bei dem Pupillendurchmesser ein signifikanter Trend von etwa 5,9 auf etwa 5,5 mm gezeigt werden, also gibt es bei diesem Messwert einen Abfall in der Versuchszeit. Bei dem PERCLOS-Wert lag in diesem Fall kein signifikantes Ergebnis vor. Von [Zha+17a] wurde ebenfalls der Pupillendurchmesser untersucht. Hier konnte auch gezeigt werden, dass der Pupillendurchmesser bei weniger Anstrengung signifikant kleiner ausfiel. In der Zusammenfassung wird erläutert, dass die Pupillenweite zwar einen Teil der Aufmerksamkeit widerspiegeln kann, aber allein nicht ausreicht, um eine Differenzierung zu ermöglichen. In dieser Arbeit wird der Pupillendurchmesser nicht berücksichtigt, da die Aufnahme dieses Messwerts meistens das Tragen eines bestimmten Sensors wie eine Brille erfordert. Dies könnte die Aufmerksamkeit des Probanden stören und ist in zukünftigen Fahrzeugen erst einmal unrealistisch umzusetzen. Eine andere Möglichkeit, die Augen auszuwerten, ist das Eye-Tracking. Hierbei wird anhand von genauen Aufnahmen der Augenbewegungen nachvollziehbar gemacht, worauf der Proband fokussiert. Ebenso werden die Längen der Fixationen ermittelt. Es wird hier also der sensorische Zustand des Fahrers erfasst. Im Kontext des automatisierten Fahrens wurden solche Daten schon von [Fel+18b] aufgenommen und ausgewertet. Hierbei wurde bei dem Vorversuch der genannten Arbeit die Anzahl aller Blicke, die mittlere Dauer eines Blickes, die kumulierte Blickdauer aller Blicke und die horizontale Blickverteilung zur Fahrszene untersucht. Diese Maße wurden alle zehn Minuten für 60 s ausgewertet. Dabei konnte festgestellt werden, dass zwischen dem Beginn der Messung und Minute 30 die Anzahl der Blicke deutlich um etwa 20 bis 40 abnahm und sich die mittlere Dauer eines Blickes von etwa 0,7 s auf etwa 1,5 s erhöhte. Dies könnte ein Indiz für Hypovigilanz sein, jedoch kann hier aufgrund der geringen Stichprobe nur

eine Tendenz gegeben werden und eine statistische Auswertung war nicht sinnvoll. Bei dem Hauptversuch wurde die prozentuale Blickzuwendung zur Fahrszene und die horizontale Blickverteilung ausgewertet. Hierbei konnte bei Probanden aus der hypovigilanten Gruppe eine signifikante Abnahme der Blickzuwendungen zwischen Minute 5 und 35 von etwa 58 % auf etwa 33 % Blickzuwendungsanteil festgestellt werden. Ein Unterschied zwischen jungen und alten Probanden konnte dabei nicht ermittelt werden. Bei der Auswertung der horizontalen Blickverteilung zeigte sich auch ein signifikanter Unterschied bei den Blicken auf die eigene Spur, die zwischen Minute 5 und 35 von 2500 auf unter 2000 Einzelbildern abnahm.

Körperpose

Die Körperpose ist vorwiegend mit dem motorischen Zustand beschrieben. Da die verschiedenen Zustände, wie zuvor erwähnt, zusammenhängen, kann davon ausgegangen werden, dass auch die Körperpose von dem kognitiven Zustand abhängig ist. Ein Zusammenhang zwischen kognitiver Belastung und verschiedenen Sitzpostionen wird beispielsweise von [Bib+19] erarbeitet. Hierbei werden sehr ähnliche Sitzdrucksensoren wie in der vorliegenden Arbeit genutzt, um die Pose auf einem Bürostuhl zu erfassen. Dabei stellt sich heraus, dass sich die Pose je nach kognitiver Belastung verändert. Es gibt verschiedene Arten die Körperpose zu bestimmen. Im Rahmen der vorliegenden Arbeit wird die Erfassung durch die erwähnten Sitzdrucksensoren stattfinden. Eine andere Möglichkeit die Körperpose zu erfassen ist beispielsweise durch eine Kameraerfassung. In einer kleineren Studie einer betreuten studentischen Arbeit hat sich herausgestellt, dass die Sitzdrucksensoren besser für die Klassifikation unterschiedlicher Posen geeignet sind, als eine Erfassung über Bildverarbeitung. Aus diesem Grund werden in der vorliegenden Arbeit nur noch die Signale der Sitzdrucksensoren als Merkmale verwendet. Die Anordnung der Sitzdrucksensoren für die vorliegende Arbeit kann der Abbildung A.3 im Anhang entnommen werden.

3

Datenverarbeitung und Fusion mittels künstlichen neuronalen Netzwerken

In diesem Kapitel geht es darum, wie die Daten aus den psychologischen Versuchen mithilfe von elektrotechnischen Methoden automatisiert verarbeitet werden können. Dabei ist die Verarbeitung auf Rohdatenebene in Abschnitt A.1.1 nachzulesen. Die Verarbeitung der resultierenden Merkmale, die Fusion und die Klassifikation werden in den folgenden Abschnitten beschrieben. Der bisherige Ansatz der Auswertung solcher psychologischen Versuche erfolgt in den meisten Fällen noch manuell von Experten. Mit einer gewissen Expertise kann zum Beispiel durch Blick auf die Zeitsignale eine Einschätzung über den Zustand des Fahrers erfolgen. Die kognitive Psychologie ist ein von [WF13] beschriebener Strang der Forschung, bei dem untersucht wird, wie biometrische Signale, so zum Beispiel das EEG, mit psychologischen Zuständen zusammenhängen. Ein analytischer Zusammenhang zwischen Körperfunktionen und Psyche lässt sich jedoch trotz fortschreitender Forschung noch nicht herstellen. Die Erkenntnisse beschränken sich in der Regel auf korrelative Zusammenhänge, die sich quantitativ belegen lassen. Aus diesem Grund ist die Art der Datenverarbeitung in der vorliegenden Arbeit größtenteils mit selbstlernenden Systemen umgesetzt. Der Vorteil hierbei ist, dass keine analytische Beschreibung für den Zusammenhang der Merkmale zum klassifizierenden Zustand bekannt sein muss.

Künstliche neuronale Netzwerke

Die Entwicklung der künstlichen neuronalen Netzwerke hat mit der Definition der McCulloch-Pitts-Zelle von [MP43] angefangen. Dieses Element ist in Anlehnung an eine biologische Nervenzelle entstanden. So werden, wie in einem biologischen System, Eingänge addiert. Ab einem bestimmten Schwellwert wird diese Zelle aktiviert (logisch wahr gesetzt), wie es auch in biologischen Nervenzellen geschieht. Eine Weiterentwicklung hierzu stellt das von [Ros58] vorgestellte Perzeptron dar, welches den elementaren Baustein in vielen künstlichen neuronalen Netzwerken bildet. Der Unterschied hier ist, dass die Eingänge mit Gewichten multipliziert werden. Diese Gewichte lassen sich anpassen. So ist es möglich, eine Vorhersage zu trainieren, indem der Soll- und Ist-Wert miteinander verglichen werden. Mit

einem einzigen solchen Element mit zwei Eingängen ist es möglich, die logischen Operationen, wie die Und-, Oder- und Nicht-Verknüpfung zu lernen und darzustellen. Die Darstellung der exklusiven Oder-Verknüpfung ist dagegen nicht mit einem einzigen Element möglich. Um dieses Problem zu lösen, werden zwei solcher Perzeptronen hintereinander geschaltet. So ergibt sich ein mehrschichtiges künstliches neuronales Netzwerk (englisch: multi layer perceptron MLP), welches weit verbreitet ist. In diesem Zusammenhang wurde von [Hor+89] gezeigt, dass ein solches mehrschichtiges Netzwerk jede beliebige Funktion abbilden kann, wenn genügend Parameter für das Netzwerk zur Verfügung stehen. Dabei ist gemeint, dass die Breite (die Anzahl der Perzeptronen in einer Schicht) und Tiefe (die Anzahl der Schichten) entsprechend groß gewählt werden muss.

Vollständig verknüpfte Netzwerke

Bei vollständig verknüpften Netzwerken (FC) ist jedes Neuron mit jedem Element der vorherigen Schicht verbunden. Es kann aus mehreren Schichten bestehen, welche unterschiedliche Anzahl an Neuronen beinhalten können. Diese Art von Netzwerken wird üblicherweise als Klassifikator verwendet. Die letzte Schicht ist dabei oftmals mit einem Ausgangsneuron pro Klasse verbunden.

softMax

Die Softmax-Funktion kann bei einer Klassifikation nach einem vollständig verknüpften Netzwerk die Wahrscheinlichkeitsverteilung der einzelnen Klassen ausgeben. Dabei muss die Summe der Klassenwahrscheinlichkeiten für jede Klasse zusammen 100 % ergeben. Dies ist hilfreich die Entscheidung zu treffen, welche Klasse von einem Netzwerk berechnet wird. So wird im Kontext der künstlichen neuronalen Netzwerke oftmals direkt die Klasse mit der höchsten Klassenwahrscheinlichkeit als einzige Ausgabe verwendet.

Long Short-Term Memory Netzwerke

Um die Zeitabhängigkeit von Merkmalen berücksichtigen zu können, wird in der vorliegenden Arbeit eine bestimmte Form von künstlichen neuronalen Netzwerken, die Long Short-Term Memory (LSTM) Netzwerke, verwendet [HS97]. Das Besondere ist, dass mit dieser Struktur eine Zeitreihe von Daten analysiert werden kann. Dies ist im Zusammenhang mit der Untersuchung von Aufmerksamkeit wichtig, da hier ein zeitabhängiger psychologischer Prozess abläuft. Gerade bei der Untersuchung von Vigilanz ist eine Zeit von mehr als 30 Minuten ausschlaggebend. Um dies zu realisieren, hat ein LSTM Element neben dem Merkmalsvektor $\mathbf{x}_t$ noch zwei weitere Eingänge, nämlich den Zellstatus

$$\mathbf{c}_t = \mathbf{v}_t \circ \mathbf{c}_{t-1} + \mathbf{e}_t \circ \mathbf{z}_t \tag{3.0.1}$$

und den versteckten Status

$$\mathbf{h}_t = \mathbf{a}_t \circ \sigma(\mathbf{c}_t). \tag{3.0.2}$$

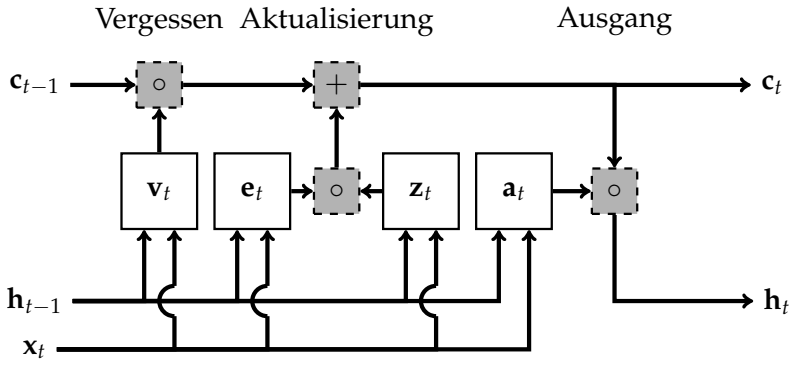


Abbildung 3.1.: LSTM Element

Der Merkmalsvektor $\mathbf{x}_t$ besteht in der hier vorliegenden Arbeit aus den Zeitreihen einzelner Merkmale, wie zum Beispiel der Zeitreihe der PERCLOS Daten eines Probanden. Dabei sind in einem LSTM Element vier Grundbausteine enthalten, wie auch in Abbildung 3.1 zu sehen ist: Eingangsgatter $\mathbf{e}_t$, Vergessensgatter $\mathbf{v}_t$, Zellkandidat $\mathbf{z}_t$ und Ausgangsgatter $\mathbf{a}_t$. Das Wort Gatter meint in diesem Fall eine Verarbeitungseinheit, welche wie eine Verengung (Gatter) aus mehreren Daten einen neuen Wert berechnet. Jeder Baustein hat eine bestimmte Funktion, die nun kurz erläutert wird. Für das Eingangsgatter kann

$$\mathbf{e}_t = \sigma_\mathrm{g}(\mathbf{w}_\mathrm{e}\mathbf{x}_t + \mathbf{r}_\mathrm{e}\mathbf{h}_{t-1} + \mathbf{b}_\mathrm{e}) \tag{3.0.3}$$

angegeben werden, das Vergessensgatter wird mit

$$\mathbf{v}_t = \sigma_\mathrm{g}(\mathbf{w}_\mathrm{v}\mathbf{x}_t + \mathbf{r}_\mathrm{v}\mathbf{h}_{t-1} + \mathbf{b}_\mathrm{v}) \tag{3.0.4}$$

bestimmt, für den Zellkandidaten gilt

$$\mathbf{z}_t = \sigma_\mathrm{z}(\mathbf{w}_\mathrm{z}\mathbf{x}_t + \mathbf{r}_\mathrm{z}\mathbf{h}_{t-1} + \mathbf{b}_\mathrm{z}) \tag{3.0.5}$$

und das Ausgangsgatter berechnet sich mit

$$\mathbf{a}_t = \sigma_\mathrm{g}(\mathbf{w}_\mathrm{a}\mathbf{x}_t + \mathbf{r}_\mathrm{a}\mathbf{h}_{t-1} + \mathbf{b}_\mathrm{a}). \tag{3.0.6}$$

Dabei ist σ_z die Aktivierungsfunktion des Zellstatus, die in der vorliegenden Arbeit als tanh gewählt wird und σ_g die Aktivierungsfunktion der Gatter, die in diesem Fall mit der gewöhnlichen Sigmoidfunktion von

$$\sigma_\mathrm{g}(x) = (1 + \mathrm{e}^{-x})^{-1} \tag{3.0.7}$$

bestimmt ist. Dazu fließen noch die anzupassenden Gewichte für den Merkmalsvektor ($\mathbf{x}_t$) $\mathbf{w}_\mathrm{e}, \mathbf{w}_\mathrm{v}, \mathbf{w}_\mathrm{z}, \mathbf{w}_\mathrm{a}$ sowie Gewichte für den versteckten Status ($\mathbf{h}_t$) $\mathbf{r}_\mathrm{e}, \mathbf{r}_\mathrm{v}, \mathbf{r}_\mathrm{z}, \mathbf{r}_\mathrm{a}$ und Bias $\mathbf{b}_\mathrm{e}, \mathbf{b}_\mathrm{v}, \mathbf{b}_\mathrm{z}, \mathbf{b}_\mathrm{a}$ in die Gleichungen ein. Die Gewichte werden bei einem Trainingsvorgang zufällig initialisiert und dann im Verlauf des Trainings anhand des Fehlers angepasst. Somit können über die beiden zusätzlichen Kanäle Informationen über die Zeit weitergegeben werden, sodass jeder Zeitschritt, aus einer Zeitreihe von Merkmalsvektoren, einen Einfluss auf die gespeicherte Information geben kann. Dabei

können in Gleichung 3.0.1 sowohl Information durch den Merkmalsvektor über das Eingangsgatter aus Gleichung 3.0.3 über eine elementweise Multiplikation mit dem vorherigen Zellstatus hinzugefügt als auch über das Vergessensgatter aus Gleichung 3.0.4 entfernt werden. Der versteckte Status, der in jedem Gatter als Eingang eine Rolle spielt, wird aus dem Zellzustand und dem Eingangsgatter wiederum mit einer elementweisen Multiplikation berechnet.

Faltende neuronale Netzwerke

Das Besondere an faltenden neuronalen Netzwerken im Gegensatz zu vollständig verknüpften Netzwerken ist, dass sich gewisse Neuronen Gewichte teilen. Sehr häufig wird diese Art von Netzwerken in der Bildverarbeitung genutzt, da mit den geteilten Gewichten eine gewisse Ortsunabhängigkeit von Merkmalen entsteht. Dabei werden oft quadratische Filter in verschiedenen Größen verwendet, um unterschiedliche Muster und Eigenschaften aus dem Bild zu extrahieren. Diese Filter werden gemäß der Faltung mit dem Eingang verrechnet. Danach werden in den meisten Fällen die resultierenden Werte mit einer sogenannten Pooling-Schicht in der Dimension verringert. Diese Art von Netzwerken soll dazu führen, dass unwichtige Störungen, wie zum Beispiel die Helligkeit des Bildes oder ein unwichtiger Hintergrund, ausgeschlossen werden. In der vorliegenden Arbeit wird diese Art von Netzwerk lediglich als Werkzeug zur Auswertung der Augen mit einer Kamera verwendet. Genaueres zu dem verwendeten Vorgehen ist Kapitel A.1.1 zu entnehmen. Die resultierenden Daten werden somit auf einer höheren Ebene weiter verarbeitet. In dieser Arbeit sollen der Nutzen dieser Daten und die Fusion mit anderen Daten im Fokus stehen.

Regression und Klassifikation mit künstlichen neuronalen Netzwerken

Die künstlichen neuronalen Netzwerke eignen sich für die Durchführung von Regressionen oder Klassifikationen. In den meisten Fällen wird hierzu ein vollständig verknüpftes neuronales Netzwerk herangezogen. Das bedeutet, dass jedes Perzeptron in dem Netzwerk mit jedem Element des Eingangsvektors in der versteckten Schicht verknüpft wird. So können die Merkmale in dem Eingangsvektor, die beispielsweise vorher von einem LSTM-Netzwerk vorverarbeitet und verdichtet wurden, ein Modell darstellen. Dieses Modell kann zwei Varianten annehmen. Eine Variante wird für die Regression verwendet und die Andere für die Klassifikation. Im Fall der Regression sind alle Perzeptronen aus der versteckten Schicht dann mit einem einzigen Perzeptron am Ausgang verknüpft. Dies kann mithilfe der Aktivierungsfunktion ein kontinuierliches oder auch diskretes Ausgangssignal erzeugen. So kann das Netzwerk eine kontinuierliche oder diskrete Funktion abbilden. Dieses Signal kann mit dem Soll-Signal verglichen werden, sodass die Gewichte in dem Netzwerk angepasst werden können. Wenn die entsprechenden Eingangssignale an dem Modell anliegen, ist es somit möglich, ein Verhalten abzubilden. Bezogen auf das Problem in der hier vorliegenden Arbeit kann somit ein Fahrerverhalten dargestellt werden. So könnte für das Beispiel der Müdigkeitserkennung das Level

der Müdigkeit kontinuierlich zwischen Null und 100 Prozent oder in Stufen, wie bei Probandenfragebögen üblich, berechnet werden. Im Fall einer Klassifikation wird nach der versteckten Schicht ein Perzeptron pro Klasse eingefügt und mit jedem Element der versteckten Schicht verknüpft. So ist es möglich, einen Wert für jede Klasse zu erhalten, mit dem eine Klassifikationswahrscheinlichkeit berechnet werden kann. Aus den verschiedenen Klassifikationswahrscheinlichkeiten kann dann eine resultierende Klasse bestimmt werden. Die prädizierte Klasse kann mit der Datenbeschriftung verglichen werden, sodass auch hier die Gewichte für die Klassifikation angepasst werden können. Bezogen auf die hier vorliegende Arbeit können so verschiedene Zustände des Fahrers erfasst werden. Da bei der Definition von Vigilanz nicht in verschiedene Stufen unterschieden, sondern eher ein Zustand beschrieben wird, den der Proband einnimmt, wird in den später vorgestellten Modellen eine Klassifikation vorgenommen. Dabei soll zwischen den sogenannten vigilanten und hypovigilanten Zuständen unterschieden werden.

Fusionstheorie

Die Fusion von Informationen kann auf verschiedenen Ebenen durchgeführt werden. Dabei wird diese von [Elm02] in drei Ebenen differenziert: Fusion auf Rohdatenebene, auf Merkmalsebene und auf Entscheidungsebene. Bei der Fusion auf Rohdatenebene werden die Sensoraufnahmen ohne eine weitere Verarbeitung mit den Daten anderer Sensoren verknüpft. Ein Beispiel für die Fusion auf Merkmalsebene ist in Abbildung 3.2 dargestellt. Die weiteren Fusionsstufen sind durch eine gestrichelte Linienführung markiert. Bei der Fusion auf Merkmalsebene werden die Merkmale unabhängig voneinander berechnet und dann vor der Klassifikation zusammengebracht. In diesem Fall gibt es nur einen Block für die Klassifikation. Sobald mehr Sensoren in der Fusion beteiligt sind, ergibt sich der Vorteil, dass mehr Merkmale in die Klassifikation einfließen können. Diese begünstigen das Training des Klassifikationsmodells, was dann in einer höheren Genauigkeit resultieren kann. Ein potentieller Nachteil besteht darin, dass die Klassifikation nur durchgeführt werden kann, wenn die Daten von allen Sensoren vorhanden und valide sind. Dazu unterschiedlich ist es bei der Fusion auf Entscheidungsebene. Hier wird, anders als in Abbildung 3.2, nicht nur ein Block als Klassifikator für die Daten aus allen Sensoren verwendet, sondern für jeden Sensor ein eigener Klassifikator verwendet. Dieses Vorgehen hat den Vorteil, dass hier das Resultat nicht abhängig ist von allen Sensoren Daten zu erhalten. Nachteilig ist allerdings, dass bei diesem Verfahren für die Klassifikation nur sehr wenige Merkmale zur Verfügung stehen. Damit könnte sich ein Training des Klassifikators erschweren und dies zu schlechteren Ergebnissen bei der Klassifikation führen.
Zudem unterscheidet [Elm02] eine weitere Kategorie der Fusion in der Art, wie verschiedene Sensoren miteinander interagieren. Dabei wird in die drei Kategorien kompetitive, komplementäre und kooperative Fusion unterschieden. Bei der kompetitiven Fusion wird mit mehreren Sensoren der gleiche Messwert aufgenommen, sodass nach der Messung beispielsweise eine Entscheidungslogik oder eine Auswahl das beste Signal auswählen kann. In dem Kontext, der vorliegenden

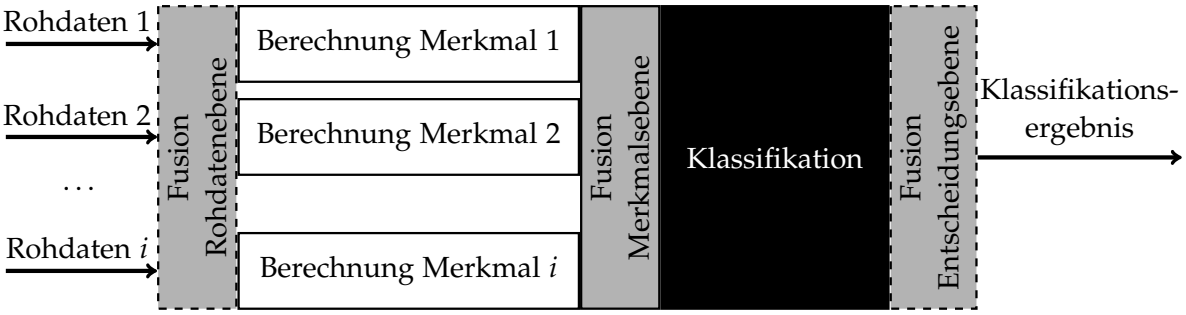


Abbildung 3.2.: Fusion auf Merkmalsebene mit Positionen für andere Fusionsebenen gestrichelt dargestellt

Arbeit, ist ein Beispiel für eine solche Fusion die Aufnahme des Herzrhythmus über zwei verschiedene Sensoren (einer optisch und einer elektrisch), welche somit dasselbe physiologische Signal erfassen. In einem zweiten Schritt könnten die Signale beider Sensoren gegenübergestellt werden und eine kompetitive Fusion durchgeführt werden. Bei der komplementären Fusion werden zwei verschiedene Umgebungen mit zwei unterschiedlichen Sensoren aufgenommen. Dies ist die einzige Möglichkeit, zwei verschiedene Umgebungen gleichzeitig zu erfassen. So werden in der hier vorliegenden Arbeit mehrere verschiedene Sensoren verwendet, die unterschiedliche physiologische Signale aufnehmen und später zu dem Gesamtbild des Fahrerzustands beitragen. Bei der kooperativen Fusion wird eine Umgebung mit zwei Sensoren zusammen erfasst. Der Unterschied zu der kompetitiven Fusion ist hier, dass beide Sensoren notwendig sind, um alle Details aus der beobachteten Umgebung zu erfassen. In der vorliegenden Arbeit findet keine kooperative Fusion statt, da jedes physiologische Signal vollständig von einem einzelnen Sensor erfasst werden kann.

Fusion von neuronalen Netzwerken

Untersuchungen zu der Fusion von neuronalen Netzwerken auf verschiedenen Fusionsstufen lassen sich bereits in der Literatur finden. Nachfolgend werden einige Beispiele angeführt, um einen Eindruck über den Stand der Technik in diesem Zusammenhang zu erlangen. Dabei lassen sich unterschiedliche Ergebnisse zu dem Vergleich zwischen den verschiedenen Fusionsstufen finden. In der Arbeit von [Gad+20] werden zwei verschiedene Fusionsmethoden diskutiert. Hierbei wird zwischen früher und später Fusion unterschieden. Bei der frühen Fusion werden die Daten vor der Klassifikationsschicht durch Aneinanderhängen fusioniert. Bei der späten Fusion wird zunächst für die Daten aus den einzelnen Domänen jeweils ein separates Netzwerk durchlaufen und anschließend durch Addition oder Mittlung der Ergebnisse fusioniert. Dabei hat für diese Anwendung die frühe Fusion besser funktioniert als die späte Fusion. Der Unterschied in der Genauigkeit liegt in diesem Fall bei vier Prozentpunkten.
In der Arbeit von [Jin+17a] werden unter anderem, wie in der hier vorliegenden Arbeit, zwei der drei Fusionsstufen ähnlich umgesetzt. Zum einen wird die Fusion der Daten aus verschiedenen Quellen (Bildern, Texte oder soziale Netzwerke) auf der Merkmalsebene durch Aneinanderhängen der Merkmale vor der Klassifikation

durchgeführt. Zum anderen wird auf Entscheidungsebene durch Mittlung der Klassenwahrscheinlichkeiten von den Klassifikatoren der einzelnen Datenquellen fusioniert. Dabei ist in der zitierten Arbeit die Gesamtgenauigkeit der verschiedenen Fusionsstufen ausgewertet worden. Es konnte festgestellt werden, dass die Fusion auf Entscheidungsebene bei einer der beiden Auswertungen bessere Resultate gezeigt hat und bei der zweiten Auswertung schlechter als die Fusion auf Merkmalsebene war. Dabei liegt der Unterschied in der ersten Auswertung bei etwa sieben Prozentpunkten und bei der zweiten Auswertung bei etwa drei Prozentpunkten.
In der Literatur ist exemplarisch ein Vergleich zwischen den drei Fusionsstufen zu finden. Hintergrund in dem Beitrag von [Jin+17b] ist die Fehleranalyse in mechanischen Komponenten mit verschiedenen Sensoren. In der Analyse, welche Fusionsmethoden sich am besten eignen, um die Daten aus den verschiedenen Domänen (akustisch, mechanisch, visuell) zu fusionieren, stellt sich heraus, dass bei der Klassifikation mit neuronalen Netzwerken die Performance mit dem Fusionslevel abnimmt. In diesem Fall ist die höchste Genauigkeit bei der Fusion auf Rohdatenebene erreicht worden. Danach folgt die Fusion auf Merkmalsebene, welche nur etwa einen halben Prozentpunkt schlechter ist. Am wenigsten Performance erreicht die Fusion auf Entscheidungsebene, welche etwa sechs Prozentpunkte schlechter ist als das beste Verfahren. Anders als in der zitierten Arbeit werden in der hier vorliegenden Arbeit die verschiedenen Modelle nicht Ende zu Ende trainiert.
Die Fusion von verschiedenen Sensoren im Kontext der Fahrerzustandserkennung wird bereits von [Jai+16] beschrieben. In diesem Fall ist das Ziel der Untersuchung jedoch keine Erkennung des Fahrerzustands, sondern die Prädiktion in die Zukunft. Dabei werden für die Berechnung der Merkmale ebenfalls wie in der hier vorliegenden Arbeit LSTM-Netzwerke verwendet. Eine wesentliche Erkenntnis aus der hier zitierten Arbeit ist, dass eine Fusion auf Rohdatenebene für diese Anwendung nicht so gut geeignet ist. Es wird beschrieben, dass bessere Ergebnisse erzielt werden können, wenn die einzelnen Sensoren ein separates LSTM-Netzwerk für die Berechnung der Merkmale nutzen. Im Zuge der zitierten Veröffentlichung wurden verschiedene Ansätze getestet. Bei dieser Untersuchung stellte sich heraus, dass die Kombination aus LSTM-Netzwerken und einem neuronalen Netzwerk als Klassifikator am besten gegen andere Ansätze wie unter anderem Support-Vector-Maschine und Random Forest abschneidet.
Eine weitere Stärke von der Fusion auf Entscheidungsebene ist eine erhöhte Ausfallsicherheit. [Zha06] untersucht diesen Zusammenhang. Hierbei wird herausgestellt, dass die Kombination von mehreren neuronalen Netzwerken für eine erhöhte Ausfallsicherheit sorgt. Dabei wird für die Fusion auf Entscheidungsebene eine modifizierte Mehrheitsabstimmung herangezogen. In diesem Fall wird ähnlich zu der hier vorliegenden Arbeit eine binäre Klassifikation verwendet. Der Unterschied besteht jedoch darin, dass in der zitierten Arbeit sehr viele gleiche Netzwerke, die auf der gleichen Datenbasis trainiert wurden, fusioniert wurden. In der vorliegenden Arbeit werden dahingegen Netzwerke auf Basis von verschiedenen Sensoren fusioniert, sodass viel weniger Netzwerke für die Fusion zur Verfügung stehen.

4

Machbarkeitsstudie zum Aufmerksamkeitsverhalten

Um einen ersten Eindruck über die Machbarkeit der Integration einer Aufmerksamkeitsaufgabe in die Fahrsituation zu bekommen, wird zunächst ein Vorversuch durchgeführt. Die dabei anfallenden Sensordaten sollen später ebenso exemplarisch dazu dienen, eine erste Fusionsmöglichkeit und Modellierung zu prüfen. In diesem Abschnitt geht es zuerst um das Design der Studie mit der dazugehörigen Fahrsimulation und Befragung. Dazu werden die Ergebnisse aus der psychologischen Befragung zum Selbstbefinden dargestellt.

4.1. Design einer Signalentdeckungsaufgabe und Aufbau der Studie

Bei dieser Vorstudie handelt es sich um einen Versuch innerhalb eines statischen Fahrsimulators. Bei diesem wird eine längere, automatisierte Autobahnfahrt simuliert und parallel dabei ist eine Signalentdeckungsaufgabe zu bewältigen. Es geht um die kontinuierliche Messung der Aufmerksamkeit während des Versuchs, ohne eine Übergabe zwischen dem automatisierten und manuellen Fahren. Wie in Kapitel 2 beschrieben, würde eine solche Situation die phasische zentralnervöse Aktivierung auslösen und so den hypovigilanten Zustand beenden. Der Aufbau besteht aus einem Bildschirm der sich frontal vor dem Probanden befindet und einer "Sitzkiste", in der zwei Autositze sowie Lenkrad und Pedalerie angebracht sind. Zudem bietet der Simulator die Möglichkeit, verschiedene Sensoren flexibel zu platzieren.
In der hier vorliegenden Arbeit wird SCANeR von AV Simulation für die Fahrsimulation benutzt. Dieser Fahrsimulator kann verschiedene Aufgaben bewältigen, wie das Erzeugen und Auslesen von CAN-Bus-Signalen und das damit verbundene Testen von Algorithmen direkt auf den Steuergeräten. Damit ist es möglich, eine Vielzahl von Sensoren und Aktoren zu simulieren und zu evaluieren. In der hier vorliegenden Arbeit beschränkt sich die Nutzung auf die Ego-Perspektive aus einem Fahrzeug. Da die Probanden nur diese Ansicht sehen, ist es nicht relevant, welches Fahrzeug simuliert wird. Es gibt keinen Bezug zu Marken oder anderen

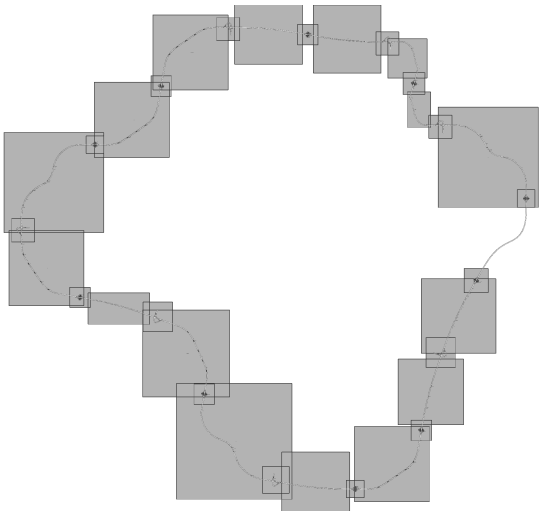


Abbildung 4.1.: Strecke von über 100 km Länge mit Start im Osten Richtung Norden

Eigenschaften des Fahrzeugs. Mit diesem Vorgehen kann ausgeschlossen werden, dass das Fahrzeug Einfluss auf Vertrauen oder Motivation der Probanden nimmt.

In der Fahrsimulation fährt das Fahrzeug auf einer über 100 km langen Autobahnstrecke mit maximal 130 $\frac{\text{km}}{\text{h}}$ auf zwei- oder dreispurigen Streckenabschnitten. Es wird versucht, eine möglichst monotone Strecke zu definieren, wie es auch von [Fel+18a] beschrieben wird. In der genannten Arbeit wurde hierfür eine absolut gerade Strecke verwendet und es gab keine Änderung in der Umgebung. Im Gegensatz dazu werden in der vorliegenden Arbeit fast nur kurvige Streckenabschnitte genutzt, weil diese bei der Fahrsimulation bereits implementiert vorliegen. Zu jedem Streckenabschnitt gehört auch eine umgebende Landschaft, was zu einer realistischeren Wahrnehmung führt. Dabei wechselt sowohl die Vegetation und Begrenzung zwischen den Fahrbahnen als auch die Umgebung links und rechts von dieser. Es gibt Streckenabschnitte über und unter Brücken, jedoch keinen Tunnel. Die umgebende Landschaft beinhaltet Gebäude, Berge und verschiedene Vegetationen. Die Gesamtstrecke setzt sich aus vier unterschiedlichen Streckenabschnitten zusammen, die auch in den verschiedenen Fahrtrichtungen befahren werden. Die Streckenabschnitte haben Längen von zehn, sieben, sechs oder vier Kilometern. Jede Streckenabschnittskombination wird durch ein Autobahnkreuz oder -dreieck verbunden. So ist nach mindestens zehn Kilometern eine Abfahr- und Auffahrsituation gegeben, wobei das Fahrzeug in keinem Fall die Autobahn verlässt oder auffährt. Es wird eine Verkehrssituation mit mittlerem Verkehrsaufkommen auf beiden Fahrbahnseiten simuliert, welche bei jeder Simulation zufällig neu im Umkreis des Fahrzeugs generiert wird. Dabei gibt es sowohl Fahrzeuge auf der eigenen Spur als auch auf anderen Spuren der gleichen Fahrtrichtung und ebenso auf der Gegenfahrbahn. Das Ego-Fahrzeug überholt langsamere Fahrzeuge durch Überholmanöver und wird auch von anderen Fahrzeugen überholt. Andere Verkehrsteilnehmer fahren auf die Autobahn oder verlassen sie an entsprechenden Auf- und Abfahrtsstellen wieder. Um die Verkehrsbedingungen für jeden Probanden gleich zu halten, wird für diesen Versuch ein Video von 45 Minuten verwendet,

Tabelle 4.1.: Synthetisch eingebrachte Schilder

plausibel	nicht plausibel
	

welches etwa dreiviertel der Runde abbildet.

Da die Vigilanz über die Veränderung von Reaktionszeiten messbar ist, wird eine Signalentdeckungsaufgabe in die Fahrsituation gebracht. In diesem Fall sollte diese nicht komplett unabhängig von der Fahraufgabe sein, wie es zum Beispiel eine auditive Nebenaufgabe wäre. Um dies umzusetzen, werden synthetische Straßenschilder am Fahrbahnrand in die Fahrsituation integriert, um eine Vigilanzaufgabe, wie in [Mac48] beschrieben, einbringen zu können. Eine ähnliche Methode wird auch von [Fel+18b] verwendet. Hierbei werden auch Schilder in die Fahrsimulation eingebettet. Jedoch weicht, der dort vorgestellte Ansatz ab, da in den Versuchsinstruktionen nicht direkt gezeigt und gesagt wurde, welche Schilder eingebracht wurden. Erst nach der Fahrt wird nach besonderen Vorkommnissen gefragt und dann eine Liste mit möglichen Schildern vorgelegt. Dabei konnte gezeigt werden, dass die Probanden nicht von sich aus die aufgetretenen Schilder wiedergeben konnten, jedoch ist es einigen Probanden gelungen, diese aus der Liste auszuwählen. Die Schilder sind, in der vorliegenden Arbeit, etwa 900 m voneinander entfernt mit einer Varianz von 90 m, um einem Gewöhnungseffekt entgegenzuwirken. Für die Vigilanzaufgabe sind zwei Arten von Signalen wichtig: zum einen die Signale, auf die nicht reagiert werden muss und zum anderen die Signale, auf die der Proband reagieren soll.

Für die Umsetzung werden zwei Arten von Schildern für die Machbarkeitsstudie verwendet: Schilder, die realistisch auf der Autobahn zu finden sind und unplausible Schilder, die nicht in einer üblichen Fahrsituation auf der Autobahn zu sehen sind. Die Schilder sind in Tabelle 4.1 abgebildet. Zu den fünf normalen Schildern gehören die Geschwindigkeitsbegrenzung von 130 $\frac{km}{h}$, der Mindestabstand von 70 m für LKW, Seitenwind, alle Verbote aufgehoben und eine Ankündigungstafel. Zu den drei unplausiblen Schildern gehören Fußgängerüberweg, Fußgängerzone und Vorfahrtstraße. Keines dieser Schilder hat Einfluss auf die Verkehrsteilnehmer. Auf die nicht plausiblen Schildern soll der Proband mit einem Tastendruck reagieren. Um die Aufgabe am Anfang leichter und später schwieriger zu gestalten, werden in der ersten Hälfte der Versuchszeit 40 % und danach nur noch 20 % der Schilder nicht plausible gewählt. Das bedeutet, dass etwa alle 25 s ein Schild am Straßenrand zu sehen ist, aber am Ende nur alle zwei bis drei Minuten auf ein Schild reagiert werden muss. Dabei wird die These aufgestellt, dass sich die Reaktionszeiten zum Ende des Fahrversuchs erhöhen. Die Aufmerksamkeitsaufgabe ist also analog zu der These aus [Mac48] gestaltet. Dazu ist sie noch auf eine Situation im Straßenverkehr bezogen. Eine Bestätigung dieser These würde die Definition des Begriffes Vigilanz entsprechen.

Um den Versuchsablauf besser nachvollziehen zu können, folgen nun Erläuterungen zu den einzelnen Schritten des Versuchs: Zuerst wurde ein Fragebogen

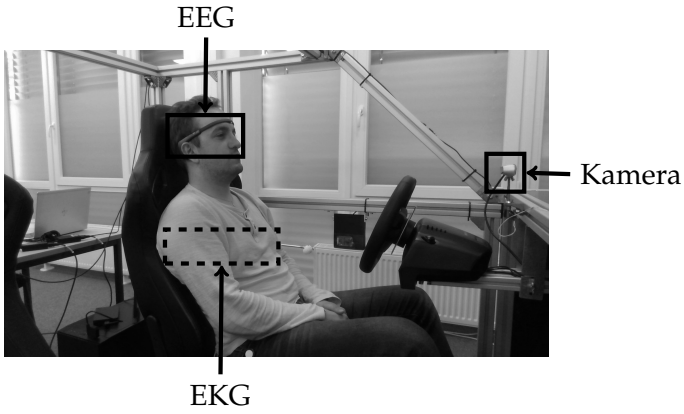


Abbildung 4.2.: Proband in dem Versuchsaufbau

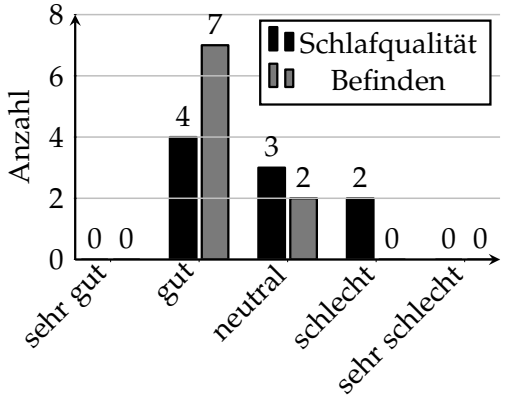


Abbildung 4.3.: Auswertung der Befragung zur Schlafqualität und Befinden vor dem Versuch

zur Person ausgefüllt, bei dem nach Alter, Geschlecht, Fahrerfahrung und aufmerksamkeitsbeeinflussenden Eigenschaften (Schlaf, Kaffeekonsum usw.) gefragt wurde. Dabei wurde für das Befinden und die Schlafqualität der letzten Nacht eine fünfstufige Skala wie folgt ausgelegt: *sehr gut*, *gut*, *neutral*, *schlecht* und *sehr schlecht*. Danach wurden die kontaktbehafteten Sensoren (EEG-Stirnband und Herz-Brustgurt) am Probanden platziert und geprüft, indem die Rohsignale und teilweise deren Varianz angezeigt wurden. Es folgten die Versuchsinstruktionen zusammen mit der Signalentdeckungsaufgabe erklärt. Nachdem der Proband, wie in Abbildung 4.2 gezeigt, in dem Fahrsimulator den Platz eingenommen hatte, wurde die Messung aller Sensoren gestartet. Dazu wurde das Video automatisch abgespielt, sodass die automatisierte Versuchsfahrt begann, wobei der Proband nur die Signalentdeckungsaufgabe zu erfüllen hatte. Nach der Versuchsfahrt verließ der Proband den statischen Fahrsimulator, legte die Sensoren wieder ab und beantwortete abschließend einen Fragebogen zur subjektiven Bewertung der Fahrt bezogen auf die erlebte Langeweile, Konzentrations-, Anstrengungs- und Müdigkeitswahrnehmung. Bei diesem Fragebogen wurde eine fünfstufige Skala wie folgt gewählt: *sehr*, *ziemlich*, *mittelmäßig*, *etwas* und *gar nicht*.

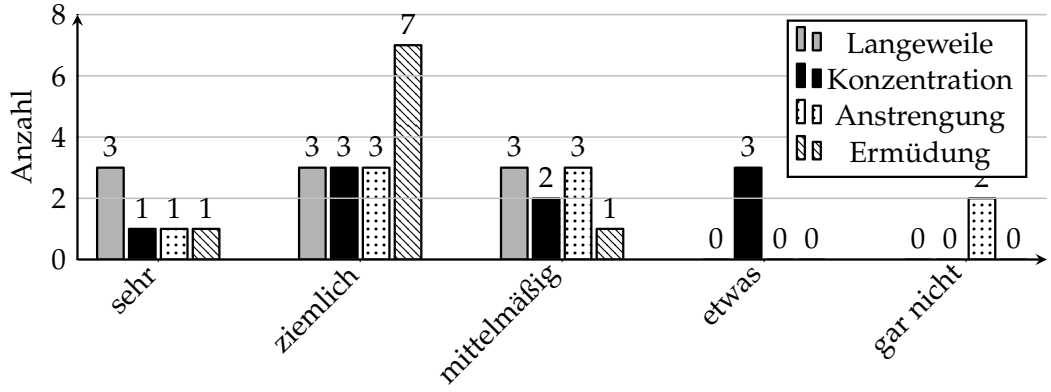


Abbildung 4.4.: Auswertung der Befragung zur Verfassung nach dem Versuch

4.2. Ergebnisse der Fragebögen und Reaktionszeiten

Nach der Beschreibung des Versuchs und dessen Ablauf, soll es in diesem Kapitel um die Auswertung aus psychologischer Sicht gehen. Es werden die Ergebnisse aus den Fragebögen und den Reaktionszeiten vorgestellt. Diese Ergebnisse sollen eine Vorstellung zu dem Hintergrund der Daten liefern, also wie sich die Situation für die Probanden angefühlt hat und wie dies auch objektiv veranschaulicht werden kann.
An der Machbarkeitsstudie haben neun männliche Probanden mit einem Alter von ∅27,7 Jahren mit einer Standardabweichung (STD) von 3,6, die seit ∅9,0 Jahren (STD 4,2) ihren Führerschein besitzen, teilgenommen. Keiner dieser Personen war an dem Versuchsdesign beteiligt. Der Versuch hat vorwiegend nachmittags stattgefunden mit einer Versuchszeit zwischen 13 und 17 Uhr. Nur zwei der Probanden haben morgens um etwa 10 bis 11 Uhr teilgenommen. Alle Probanden besaßen mindestens die Führerscheinklasse B. In der Gruppe der Probanden haben sechs Personen eine Sehhilfe genutzt, von denen eine Person Kontaktlinsen und der Rest eine Brille trug. Aufgrund des Einflusses einer Brille auf die optische Erfassung der Augen, können diese Angaben für die folgende Datenerarbeitung von Relevanz sein. Bei Kontaktlinsen kann die Blinzel-Häufigkeit verändert sein. Zum einen können Kontaktlinsen bei einer Reizung der Augen zu einer erhöhten Blinzel-Häufigkeit führen und zum anderen können diese ebenso dafür sorgen, dass sich die Anzahl der Augenschließungen verringert. Die Schlafzeit in der Nacht vor dem Versuch lag bei ∅6,89 Stunden (STD 0,42 Stunden). Die Ergebnisse zu der Schlafqualität und dem aktuellen Befinden sind in Abbildung 4.3 dokumentiert. Dabei wird deutlich, dass die Schlafqualität vor dem Versuch nur bei etwa 44 % der Probanden als *gut* empfunden wurde und 22 % der Probanden den Schlaf als *schlecht* bewerteten. Das Befinden vor Antritt der Fahrt ist bei 78 % der Probanden als *gut* angegeben worden und der Rest der Gruppe beurteilt das Befinden als *neutral*. Mit diesen Fragen wird aber deutlich, dass es keine besonderen circadianen Einflüsse auf die Probanden geben sollte. Die Schlafzeit scheint ausreichend groß, sodass nicht von einer starken Übermüdung auszugehen ist. Die Angabe zur Schlafqualität könnte für spätere Auswertungen der physiologischen Daten ein Indiz für eine Unauffälligkeit geben. Eine generelle statistische Aussage ist bei der geringen Anzahl von Teilnehmern jedoch nicht sinnvoll.

Die Befragung nach der Fahrt wird in Abbildung 4.4 verdeutlicht. Hier wird deutlich, dass bei allen Probanden Langeweile aufgetreten war. Ein Drittel bewertete das Auftreten mit *sehr* stark, ein Drittel mit *ziemlich* stark und ein Drittel mit *mittelmäßig* stark. Die Konzentrationseinschätzung ist fast symmetrisch um die Angabe *mittelmäßig* verteilt, bis auf einen Probanden, der sich *Sehr* stark konzentrieren musste, um die Signalentdeckungsaufgabe zu bewältigen. Wird die Anstrengungswahrnehmung beobachtet, war diese von drei Probanden mit *mittelmäßig* eingestuft worden, drei Probanden schätzten diese auf *ziemlich*, eine Person auf *sehr* und zwei Probanden spürten die Anstrengung *gar nicht* während des Versuchs. Müdigkeit trat bei allen Probanden in der subjektiven Bewertung auf. Dabei schätzten sich selber 78 % der Gruppe mit *ziemlich* ermüded, 11 % mit *mittelmäßig* ermüdet und weitere 11 % mit *sehr* ermüdet ein.

Als Referenz für die Reaktionszeiten wurde das Video einmal manuell ausgewertet. Hierzu wurde das Video in Einzelbilder separiert. Diese Einzelbilder wurden dann nacheinander angesehen. Dabei wurden die Nummern der Einzelbilder festgehalten, auf denen ein unplausibles Schild gerade von einem normalen Schild zu unterscheiden war.

Um die Reaktionszeiten auf die festgehaltenen Einzelbilder bestimmen zu können, muss das Signal vom Tastendruck mit dem Video synchronisiert werden. Der Tastendruck bekommt einen Zeitstempel der aktuellen Systemzeit. Da der Videoplayer immer unterschiedlich lange zum Laden des Videos benötigt, muss der genaue Start des Videos bestimmt werden. Nur so kann ermittelt werden, was der Proband zum Zeitpunkt des Drückens gesehen hat. Hierzu wurden am Anfang des Versuches Bildschirmfotos mit Zeitstempeln der Systemzeit abgespeichert. In der Auswertung wurde dann mittels Bildvergleich automatisiert ein passendes Einzelbild aus dem Video zu dem jeweiligen Bildschirmfoto gefunden. Da bekannt ist, welche Nummer jene Einzelbilder haben, ist es möglich, für jede Bildschirmaufnahme einen Startzeitpunkt für das Video zu bestimmen. Hierzu wurde die Annahme getroffen, dass das Video mit 30 Bildern pro Sekunde abgespielt wird. Durch diese Information kann mit der Nummer des Einzelbildes und dem Zeitpunkt der Bildschirmaufnahme die Startzeit des Videos in der Systemzeit berechnet werden, die auch für die Aufnahme der Tastendrücke verwendet wurde. Um die Reaktionszeiten zu bestimmen, wurde ebenfalls die Annahme getroffen, dass die Einzelbilder regelmäßig mit 30 Bildern pro Sekunde abgespielt werden. Da die Nummern der Einzelbilder, wie oben beschrieben, festgehalten worden sind und die Startzeit des Videos bekannt ist, kann hiermit eine Zeit ausgerechnet werden. Diese Zeit wird dann mit dem Zeitstempel des Tastendrucks verglichen, um die Reaktionszeit zu bestimmen.

In Abbildung 4.5 sind die durchschnittlichen Reaktionszeiten mit den zugehörigen Standardabweichungen dargestellt. Zu erkennen ist, dass es mit fortdauerndem Versuch einen Trend hin zu kleineren Reaktionszeiten gibt. Es wird darauf hingewiesen, dass eine statistische Auswertung bei der geringen Anzahl von neun Probanden nicht sinnvoll ist. Aus diesem Grund kann keine Aussage über die Signifikanz des Trends oder den Einfluss der Variable Versuchszeit getroffen werden. Eine negative Reaktionszeit ist dabei möglich, da der Zeitpunkt des visuellen Reizes

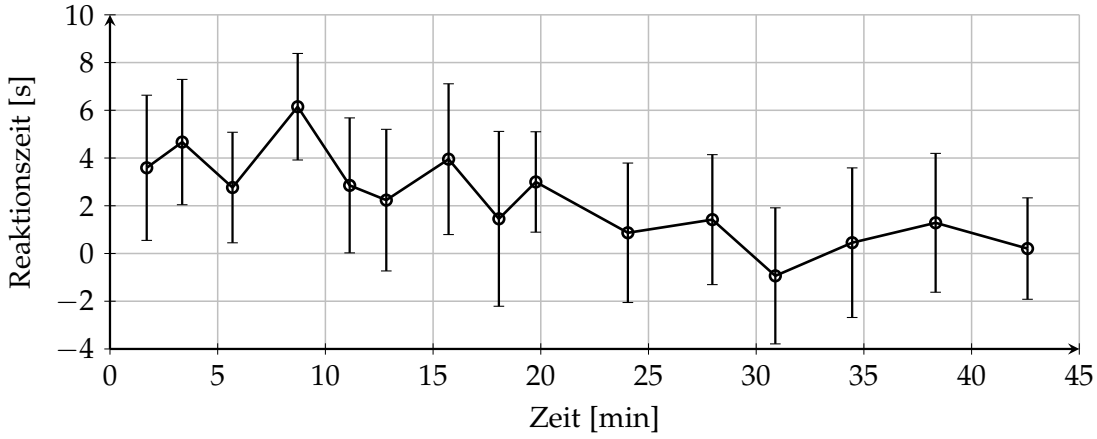


Abbildung 4.5.: Reaktionszeiten der Probanden auf die Straßenschilder

auf Grundlage von Einzelbildern bestimmt wird, auf denen das Straßenschild klar von anderen unterschieden werden kann und die Probanden vor diesem Zeitpunkt den Knopf drücken können. Die Probanden scheinen das Erkennen der Schilder durch Merkmale wie Farbe oder Größe zu lernen und sind so in der Lage, die Schilder schneller voneinander zu unterscheiden. Zudem könnte es auch einen Unterschied zwischen einem Video und der Auswertung von Einzelbildern geben. Durch das Video ist ein Schild durch die Bewegung möglicherweise schon eher zu erkennen. Um diesem Phänomen bei zukünftigen Arbeiten entgegenzuwirken, könnte ein Versuch durchgeführt werden, bei dem eine hohe Reizrate nur mit unplausiblen Schildern vorliegt. So könnte die durchschnittliche Reaktionszeit für unplausible Schilder bei vigilanten Probanden ermittelt werden, welche dann als Referenz dienen könnte.
In der vorliegenden Arbeit wurde jedoch untersucht, ob der Trend des Lerneffekts länger anhält. Es werden hierfür zufällig zwei Probanden aus der Gruppe der Teilnehmenden erneut mit der doppelten Versuchszeit getestet, wobei diesen zweimal die gleiche Fahrt direkt hintereinander abgespielt wird. Die Reaktionszeiten der ersten Person sind in Abbildung 4.6 dargestellt. Hier wird deutlich, dass die Reaktionszeiten in der zweiten Hälfte des Versuchs ansteigen und deutlich größer werden. So ist die maximale Reaktionszeit nahezu doppelt so groß und steigt bis auf 8,76 s an. Dazu ist der Verlauf mit kleineren und größeren Reaktionszeiten in der zweiten Runde wiederzuerkennen. Bei dieser Testperson steigen jedoch die Reaktionszeiten in der zweiten Runde um ∅3,7 s an. In Abbildung 4.7 ist das Ergebnis des zweiten Probanden abgebildet. Hier wird deutlich, dass die Reaktionszeiten in der zweiten Hälfte des Versuchs nicht besonders stark ansteigen. Zwar liegt in der zweiten Hälfte die maximale Reaktionszeit mit 11,41 s vor, jedoch setzt sich dieser Trend nicht bis zum Ende hin fort. In Minute 24 der zweiten Runde wird eine Reaktionszeit von etwa -2,06 s gemessen, die sich somit nur um 0,77 s von der Reaktionszeit aus der ersten Runde unterscheidet. Zudem wird deutlich, dass sich die Reaktionszeiten am Ende des Versuchs ab Minute 24 innerhalb der Standardabweichung der Reaktionszeiten der ersten Runde aus Abbildung 4.5 befinden. Dazu kann noch festgehalten werden, dass bei dem zweiten Probanden die Varianz der Reaktionszeiten zunimmt. Dies könnte wiederum für eine erhöhte Hypovigilanz sprechen. Es könnte sein, dass es

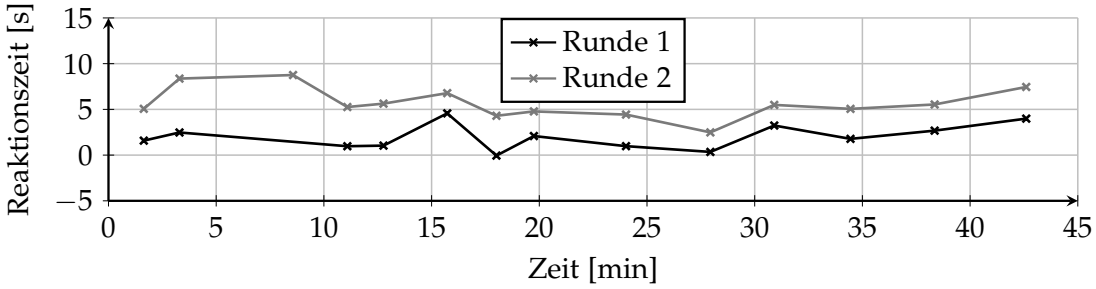


Abbildung 4.6.: Reaktionszeiten der ersten trainierten Person bei längerer Fahrt

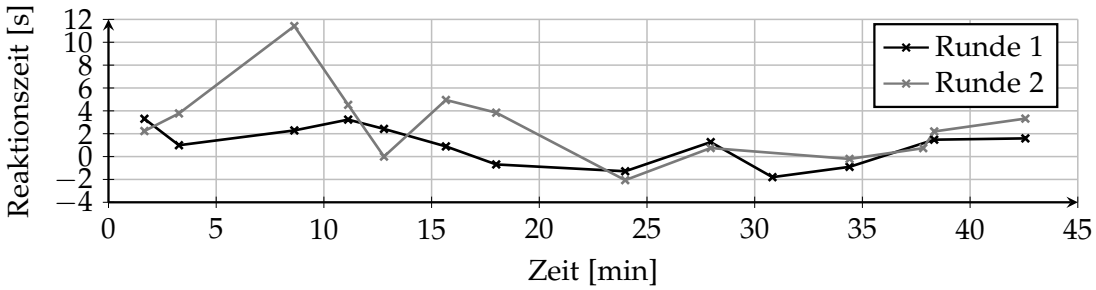


Abbildung 4.7.: Reaktionszeiten der zweiten trainierten Person bei längerer Fahrt

für den Probanden an einigen Zeitpunkten schwieriger war, die Aufmerksamkeit auf einem hohen Niveau zu halten. Allerdings nimmt diese erhöhte Varianz zum Ende des Versuches wieder ab und die Reaktionen sind nahezu identisch.
Insgesamt muss festgehalten werden, dass die in diesem Kapitel vorgestellte Aufmerksamkeitsaufgabe nicht alleine als Vigilanzaufgabe geeignet ist. Die Reaktionszeiten zeigen nicht das typische Verhalten von abnehmender Vigilanz. Die Aufgabe könnte zu simpel sein, sodass sich die volle Aufmerksamkeit auf die Vigilanzaufgabe richtet und so ein besseres Ergebnis erzielt werden kann. Somit werden die Reaktionszeiten nicht zur Einordnung der physiologischen Daten genutzt, sondern die Gehirnaktivität hierfür herangezogen.

5

Probanden-Studie zur Vigilanzuntersuchung im Fahrsimulator

In diesem Kapitel wird das Vorgehen zum Hauptversuch erläutert. Im Mittelpunkt steht dabei die Betrachtung des automatisierten Fahrens selbst als Vigilanzaufgabe. Damit soll das Verhalten der Probanden in diesen Situationen untersucht werden. Darüber hinaus soll gezeigt werden, dass verminderte Vigilanz und andere aufmerksamkeitsbezogene Konstrukte ausgelöst werden. Mit Verhalten ist in diesem Fall die physiologische Reaktion der Probanden auf diese Aufgabe gemeint. Die resultierenden Daten aus dem Versuch sollen später als Grundlage für das datenbasierte Modell zur Klassifizierung der Aufmerksamkeit des Fahrers dienen.

5.1. Design einer Vigilanzaufgabe mit unterschiedlichen Reizen

Der Vorversuch hat gezeigt, dass eine Signalentdeckungsaufgabe, wie das Reagieren auf unplausible Schilder, vermutlich zu einfach ist. Um nicht, wie in Kapitel 2, die selektive Aufmerksamkeit anzusprechen, sondern wie bei dem manuellen Fahren in den Bereich einer geteilten Aufmerksamkeitsaufgabe zu kommen, wurden in diesem Experiment mehrere verschiedene Elemente in die Fahrsituation eingebracht. Hierbei wurde versucht, die Überwachungsaufgabe für ein Fahren unter SAE Level 2 abzubilden. Hierbei muss der Fahrer in jedem Fall unsichere Situationen erkennen und auf diese reagieren. In dem Szenario dieser Studie sind drei verschiedene Situationen in der Simulation abgebildet, welche allesamt gefährlich für das Fahrzeug und den Fahrer werden könnten. Diese Situationen sollen als Reize dienen. Der Proband soll so schnell wie möglich per Tastendruck bestätigen, dass er das Fahrzeug übernehmen würde. Eine Übergabe wird in diesem Fall nicht vollzogen und das Fahrzeug fährt nach dieser Situation unfallfrei weiter. Ein Hindernis auf der Straße ist eine dieser Situationen. Diese haben in keinem Fall Einfluss auf das Fahrzeug oder die Simulation und es tritt lediglich ein sichtbares Objekt auf. Zwei verschiedene Elemente wurden eingebracht, zum einen ein Quader mit grauen Steinen als Textur und zum anderen ein zylinderförmiges Fass. Der Quader hat eine Grundfläche

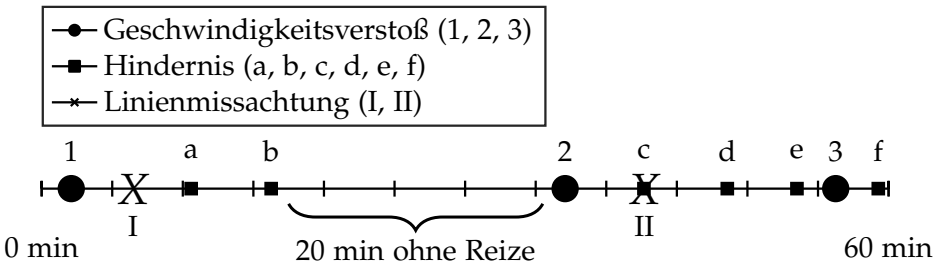


Abbildung 5.1.: Zeitstrahl mit verschiedenen Reizen

von 1 m mal 1 m und eine Höhe von 0,2 m, damit er schwieriger von der Straße zu unterscheiden ist. Das Fass erscheint in der Fahrsimulation grau und ist ebenfalls flach auf der Straße. Eine weitere Situation, welche als unsicher gilt, ist das Überfahren von Straßenmarkierungen. In diesem Fall wird auf einer zweispurigen Straße die Spurtrennung mittels durchgezogener Linie markiert, wodurch verkehrsrechtlich kein Spurwechsel vollzogen werden darf. Durch ein langsam fahrendes Fahrzeug auf der gleichen Spur des Ego-Fahrzeugs wird ein automatisierter Spurwechsel ausgeführt. Somit missachtet das automatisierte Ego-Fahrzeug, in dem der Proband in der Simulation mitfährt, die durchgezogene Linie. Eine andere Situation wird durch die Geschwindigkeitsbegrenzung gegeben. Hierbei gibt es entlang der gesamten Strecke verschiedene Geschwindigkeitsbegrenzungsschilder. Diese sind, anders als in der Pilotstudie, alle auf der Autobahn zu finden. Zu den Schildern gehören: Geschwindigkeitsbegrenzung von 90 $\frac{\text{km}}{\text{h}}$, 110 $\frac{\text{km}}{\text{h}}$, 130 $\frac{\text{km}}{\text{h}}$, *Alle Verbote aufgehoben* und *Achtung Seitenwind*. Die Schilder zur Geschwindigkeitsbegrenzung sowie *alle Verbote aufgehoben*, haben in dieser Studie Einfluss auf das Fahrzeug. Die Automatisierung in dem Fahrzeug passt sich entsprechend an die Begrenzung an oder hebt die Begrenzung auf und fährt die gültige Richtgeschwindigkeit auf der Autobahn in Deutschland von 130 $\frac{\text{km}}{\text{h}}$. An der Stelle von drei verschiedenen Schildern missachtet das automatisierte Fahrzeug die Geschwindigkeitsbegrenzung. Nach der Geschwindigkeitsbegrenzung beschleunigt es oder bremst nicht auf die Geschwindigkeitsbegrenzung ab.

Innerhalb des Versuches treten diese unsicheren Situationen mehrmals auf. In Abbildung 5.1 ist die Abfolge dargestellt. Jeder Teilstrich trennt hierbei einen Abschnitt von fünf Minuten. Jede Situation tritt in den ersten 15 Minuten einmal auf, sodass davon ausgegangen werden kann, dass der Proband jedes Signal (1,I,a) einmal in einem aufmerksamen Zustand erlebt. Nach dem Hindernissignal b gibt es eine Pause von 20 Minuten, in der keine Reizsignale auftreten. Diese soll dafür sorgen, dass die Probanden ermüden und die Reizrate selten wird. In dieser Pause sind dennoch verschiedene Geschwindigkeitsbegrenzungen, die in diesem Fall alle korrekt von dem automatisierten Fahrzeug abgefahren werden. In Minute 42 gibt es zwei kurz aufeinander folgende Reize. Hierbei wechselt das Fahrzeug von der rechten Spur auf die linke Spur, auf der sich dann das Hindernis befindet. Für die Untersuchung der Vigilanz sind ausbleibende oder verzögerte Reaktionen von Interesse.

Da später in der vorliegenden Arbeit die Geschwindigkeitsverstöße eine erhöhte Bedeutung erhalten, werden diese Reize noch etwas genauer beschrieben. Bei dem

Reiz 1 handelt es sich um ein Geschwindigkeitsbegrenzungsschild mit dem Wert 90 $\frac{km}{h}$, welches an dieser Stelle nun missachtet wird. Das Fahrzeug beschleunigt bis auf 130 $\frac{km}{h}$, anstatt der Begrenzung von 90 $\frac{km}{h}$ zu folgen. Reiz 2 ist sehr ähnlich, denn hierbei ist auch das Schild Geschwindigkeitsbegrenzung 90 $\frac{km}{h}$ an dem Zeitpunkt zu sehen. In diesem Fall fährt das Fahrzeug wegen dem vorherigen Schild mit einer Geschwindigkeit von 110 $\frac{km}{h}$. Diese müsste an diesem Schild auf 90 $\frac{km}{h}$ verringert werden. Das Fahrzeug verzögert jedoch nicht, sondern fährt unverändert mit 110 $\frac{km}{h}$ weiter. Das Schild *Seitenwind* ist bei Reiz 3 von Bedeutung, was ihn von den anderen unterscheidet. Aus den vorherigen Schildern geht dieses Mal die Geschwindigkeitsbegrenzung von 110 $\frac{km}{h}$ hervor. Diese sollte sich durch das Schild *Achtung Seitenwind* normalerweise nicht ändern. Das automatisierte Fahrzeug beschleunigt allerdings an dieser Stelle bis auf 130 $\frac{km}{h}$, ähnlich zu dem Verhalten bei Reiz 1.

Fragebögen

Bereits bei der Machbarkeitsstudie, welche in Kapitel 4 erläutert wurde, konnten ersten Erfahrungen mit Fragebögen gesammelt werden. Hier wurde nach gewissen subjektiven Empfindungen, wie Langeweile, Konzentration, Anstrengung und Ermüdung gefragt. Ein wesentliches Problem bei diesem Versuch war jedoch, dass lediglich nach der Fahrt nach diesen speziellen Empfindungen gefragt wurde und vor dem Versuch nur allgemein nach dem Befinden. Dies führt dazu, dass die Veränderung der Empfindung durch den Versuch nicht deutlich wird. Ein Vergleich der subjektiven Einschätzung des Probanden zwischen vor und nach dem Versuch kann somit nicht durchgeführt werden. Eine weitere identifizierte Schwachstelle der Befragung vor dem Versuch ist die fehlende Einordnung der Skala. Da die genutzten Fragen nicht auf einem aus der Literatur bekanntem Beispiel basieren, fällt es schwer, die Resultate einzuordnen. Ein besserer Ansatz ist bekannte Skalen für die Messung der subjektiven Einschätzung zu nutzen. Diese Erkenntnisse werden nun berücksichtigt.
Um neben den Reaktionszeiten auch eine subjektive Einschätzung zum Verhalten der Probanden zu erlangen, füllen diese mehrere Fragebögen aus. Zunächst wurde ein Personenfragebogen genutzt, bei dem nach Alter, Geschlecht, Gewicht, Größe, Fahrleistung, Sehhilfen, Schlafzeit, Kaffeekonsum, Gehirnerschütterung, aufmerksamkeitsbeinflussenden Medikamenten und Bluthochdruck gefragt wird. Mit der Größe und dem Gewicht sollen Informationen gegeben werden, die Probleme bei den Signalen der Sitzdrucksensoren aufdecken können. Fragen zu Sehhilfen, Gehirnerschütterungen und Blutdruck sollen dazu beitragen, ungewöhnliche Messungen der physiologischen Daten einordnen zu können. So dient die Information über Gehirnerschütterungen als Ausschlusskriterium, da sich hiermit die Gehirnaktivität anders verhalten kann und so nicht mehr eingeordnet werden könnte. Die Fahrleistung soll einen Einblick darüber geben, ob Probanden längere Fahrten gewöhnt sind oder eher nicht.
Zudem wird, ähnlich wie in [Fel+18b] beschrieben, ein *Mind Wandering Questionnaire* verwendet, um die Ablenkungsbereitschaft der Probanden besser einschätzen

zu können. Erfasst werden soll deshalb, ob eine Person zu Tagträumen oder Sekundenschlaf neigt. Anders als in [Kör+15], wird hierbei nicht ein Auszug aus dem *Dundee Stress State Questionnaire* (DSSQ), sondern eine Kurzversion der *Short Stress State Questionnaire* (SSSQ) verwendet. Der von [HN15] vorgestellte Fragebogen enthält 24 Fragen. Diese sind für die hier vorliegende Arbeit in das Deutsche übersetzt worden. Die 24 Fragen werden einmal vor dem Versuch und leicht abgewandelt nach dem Versuch gestellt. Die Fragen eins bis zehn zielen auf den Gefühlszustand des Fahrers ab. Dieser hat Einfluss auf die Motivation des Probanden, welche wiederum Einfluss auf die Vigilanz haben kann. Dazu gibt es weitere Fragen (11.-24.) konkret zur Motivation und zu möglichen Ablenkungen, die bei einer Unterforderung auftreten können. Ebenso wird eine Einschätzung über die Erwartung an den Versuch abgefragt. Jede der 24 Fragen wird auf einer fünfstufigen Skala eingeordnet, die immer mit den Worten *gar nicht*, *ein wenig*, *etwas*, *sehr* und *äußerst* beschriftet ist. Vor dem Versuch wird nach Möglichkeit das aktuelle Befinden erfragt. Generellen Eigenschaften sollten hier nicht ermittelt werden, da sich diese über den Versuch nicht ändern sollten. Nach der simulierten Fahrt sind die Fragen auf den Zeitraum des Versuchs bezogen. Eine beispielhafte Frage ist, ob der Proband während der Fahrt zu Tagträumen geneigt war.
Zusätzlich wurde vor und nach dem Versuch noch die Müdigkeit von den Probanden auf der *Karolinska Sleepiness Scale* (KSS) eingeordnet. Diese wird von [ÅG90] vorgestellt, wobei nur einige der neun Stufen mit einer Beschriftung versehen sind. In der Arbeit von [Ake+14] werden die fehlenden Beschriftungen der Skala ergänzt, sodass alle Punkte der Skala beschriftet sind. In der hier vorliegenden Arbeit wird eine deutsche Fassung dieser Skala verwendet.
Um eine bessere Einschätzung zu bekommen, wann gewisse psychologische Zustände bei den Probanden eintraten, wurden hierzu noch sechs weitere Fragen gestellt. Falls einer der Zustände Konzentrationsverlust, Unaufmerksamkeit, Ablenkung, Müdigkeit Sekundenschlaf oder Tagtraum eingetreten war, sollte der Proband einen oder mehrere Zeitpunkte mit einem vertikalen Strich auf einem Zeitstrahl markieren. Dieser war mit einer Länge von 10 cm immer gleich lang und links mit *Anfang* sowie rechts mit *Ende* beschriftet. Zur Auswertung konnte dann später die Entfernung zwischen dem Anfang des Zeitstrahls und den Markierungen gemessen werden, sodass ein Zeitpunkt in Bezug zur Versuchszeit ermittelt werden konnte.
Die letzten drei genannten Fragebögen sind im Anhang A.2 zu finden.

Versuchsablauf

Um den Ablauf des Versuchs besser verstehen zu können, wird in diesem Abschnitt der Versuch schrittweise erklärt. Dazu kann in Abbildung 5.2 auch eine zusammengefasste Darstellung herangezogen werden. Zunächst wurde der Proband begrüßt und hatte die Möglichkeit, seine privaten Gegenstände und seine Jacke an einer Garderobe abzulegen. Anschließend wurde er gebeten den Brustgurt zur Messung des Herzens anzulegen. Ist dies erfolgt, konnte der Proband im Fahrsimulator-Aufbau Platz nehmen. Als nächstes sollte der Proband die Einverständniserklärung unterzeichnen und den Fragebogen zur Person ausfüllen.

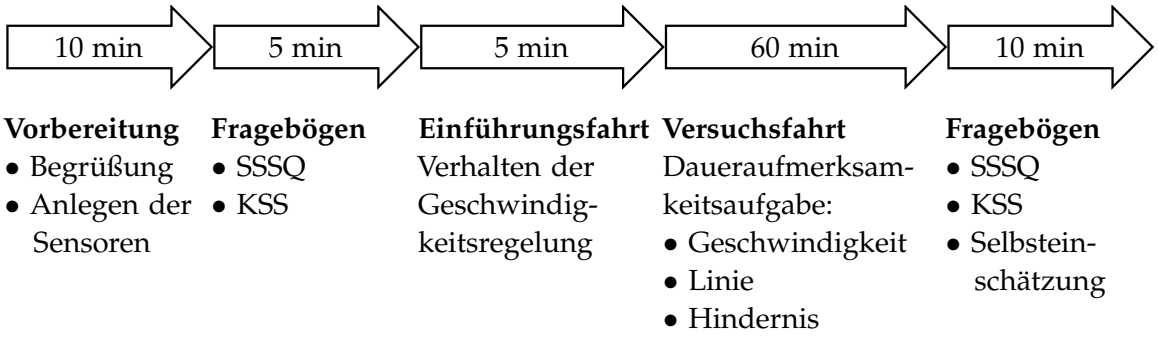


Abbildung 5.2.: Ablaufdiagramm Probandenstudie

Danach wurde der Versuch mit Hilfe der Versuchsinstruktionen erklärt. Hierbei wurde auf den Hintergrund der Untersuchung hingewiesen, erklärt, welche Sensoren ein Signal aufzeichnen und welches Verhalten eine Aufnahme stören könnte. Daraufhin wird die Signalentdeckungsaufgabe erläutert, indem auf jede gefährliche Situation eingegangen wird. Hier wird zwar die Situation beschrieben, aber nicht genau erklärt, welche optischen Signale ausschlaggebend sind. Im Anschluss wird klar gemacht, wie der Proband reagieren sollte. In diesem Fall soll so schnell wie möglich ein Knopf am Lenkrad gedrückt werden, sobald der Proband das Fahrzeug aufgrund einer unsicheren Situation übernehmen würde. Dazu wird darauf hingewiesen, dass es keine Rückmeldung an den Fahrer gibt und das Fahrzeug weiterhin die automatisierte Fahrt fortsetzt. Auch ein Fehlverhalten wird nicht durch den Knopfdruck aufgelöst, sodass hier auch explizit erklärt wird, dass das Fahrzeug durch die Hindernisse fährt. Es wird versichert, dass hierbei weder ein Nachteil für das Fahrzeug noch für den Probanden entsteht. Zuletzt wird noch auf die Versuchszeit eingegangen und welche Risiken bei dem Versuch auftreten können. Nach der Erläuterung der Versuchsinstruktionen wird mit einem Beispiel fortgefahren. Hierbei wird die Simulation gestartet und das normale Verhalten des Fahrzeugs beschrieben. Dies ist besonders wichtig, da das Fahrzeug nicht direkt bei einem Geschwindigkeitsbegrenzungsschild die gegebene Zielgeschwindigkeit erreicht hat, sondern erst einige Sekunden nach dem Schild angepasst wird. So soll vermieden werden, dass die Probanden fehlerhafte Reaktionen auf die normale Geschwindigkeitsanpassung ausführen. Allerdings wird in diesem Beispiel kein Reiz gezeigt, um keinen Trainingseffekt zu erzeugen. Nach dem Beispiel wird der SSSQ an den Probanden gegeben. Dieser sollte erst nach dem Beispiel ausgefüllt werden, da es Fragen gibt, die sich auf die Aufgabe beziehen. So kann der Proband seine Motivation für die Aufgabe besser einschätzen, wenn diese bereits erklärt wurde. Nach dem Ausfüllen des SSSQ wird noch die Müdigkeit auf der KSS abgefragt. Die Messung der Sensoren und die Simulation werden daraufhin gestartet. Während der Versuchsfahrt wird der Proband gebeten, keinen Kontakt zu dem Versuchsleiter aufzunehmen, solange keine Probleme auftreten. Der Versuchsleiter achtete während der Versuchsfahrt auf den in der Umgebung generierten Verkehr. Hier gab es Probleme, da öfter Fahrzeuge zu nah aneinander generiert wurden und dann mit stark reduzierter Geschwindigkeit gefahren sind oder stehen blieben. Diese Fahrzeuge wurden teilweise automatisiert und teilweise von dem Versuchsleiter aus

der Fahrsimulation entfernt. Da dies nicht immer rechtzeitig gelungen ist, gab es bei einigen Probanden Zusammenstöße mit anderen Verkehrsteilnehmern. Nach der Versuchszeit dunkelte der Bildschirm ab, die Messung der Sensoren wurde gestoppt. Danach wurde die Müdigkeit nach dem Versuch auf der KSS von dem Probanden eingeschätzt. Nachfolgend musste dieser noch den zweiten Teil des SSSQ ausfüllen und die Selbsteinschätzung auf den Zeitstrahlen angeben. Anschließend konnten die Sensoren abgelegt werden und der Versuch war beendet. Insgesamt hat der Versuch etwa eineinhalb Stunden gedauert. Davon wurden etwa 20 Minuten für die Aufgaben vor dem Versuch in Anspruch genommen, 60 Minuten für die simulierte Fahrt und etwa 10 Minuten für den Abschluss des Versuchs eingeplant.

5.2. Psychologische Ergebnisse der Studie

In diesem Abschnitt werden die untersuchte Gruppe von Probanden vorgestellt und die Ergebnisse aus der Messung der Reaktionszeiten sowie den Fragebögen behandelt. Auf die Auswertung der physiologischen Daten wird später in Kapitel 6 eingegangen.

Untersuchte Gruppe von Probanden

Die in der hier vorliegenden Arbeit untersuchte Gruppe besteht aus 32 Probanden, die sich in 29 männliche und lediglich drei weibliche Probanden aufteilen lässt. Eine Unterscheidung durch das Merkmal Geschlecht ist in diesem Fall also durch das Ungleichgewicht nicht sinnvoll. Dazu zeigen sich keine wesentliche Unterschiede in den Messungen der physiologischen Daten. Ein Proband war vorher als Testproband unter sehr ähnlichen Bedingungen getestet worden, allerdings wurde aus der Erfahrung dieses Pilotversuchs die Versuchsinstruktion angepasst. Es wurden gezielt Personen aus zwei verschiedenen Gruppen für den Versuch ausgewählt, die sich durch das Merkmal Alter unterscheiden. Die eine Gruppe beinhaltet dabei Probanden in einem Alter unter 30 Jahren. Die andere Gruppe war durch ein Alter von mehr als 49 Jahren charakterisiert. Wie in Tabelle 5.1 zu sehen ist, liegt der Altersdurchschnitt insgesamt bei 32,9 Jahren. Bei der jüngeren Gruppe ergibt sich ein Altersdurchschnitt von 24,3 Jahren, mit einem Mindestalter von 18 Jahren, da ein Führerschein für diesen Versuch als Voraussetzung galt. Bei der älteren Gruppe liegt der Durchschnitt bei 55,1 Jahren. Hierbei wurden überwiegend Personen ausgesucht, die über 50 Jahre alt waren, um einen größeren Unterschied zu der jüngeren Gruppe feststellen zu können. Die zwei Gruppen unterscheiden sich neben dem Alter auch in der Fahrerfahrung, wie es in der zweiten Spalte der Tabelle zu sehen ist. Hierbei geben die älteren Probanden im Durchschnitt etwa 5000 km mehr Fahrleistung pro Jahr an. In Bezug auf den der Studie vorangegangenen Schlaf unterscheiden sich die beiden Gruppen nicht. Bei dem Konsum von Kaffee entsteht wieder ein Unterschied, denn hier geben die älteren Probanden im Schnitt mit 1,8 Kaffetassen etwa doppelt so viel an wie die Gruppe der jüngeren Probanden.
In Abbildung 5.3 wird durch das Verhältnis zwischen jüngeren und älteren Probanden in dem linken Kreisdiagramm deutlich, warum das durchschnittliche Alter über alle

Tabelle 5.1.: Durchschnittliche Selbsteinschätzung zur Person

	Alter (Jahre)	Fahrerfahrung ($\frac{km}{Jahr}$)	Schlaf (Stunden)	Kaffeetassen
Gesamt	32,9	11 841	6,7	1,2
18-30	24,3	10 517	6,8	0,9
49-63	55,1	15 222	6,6	1,8

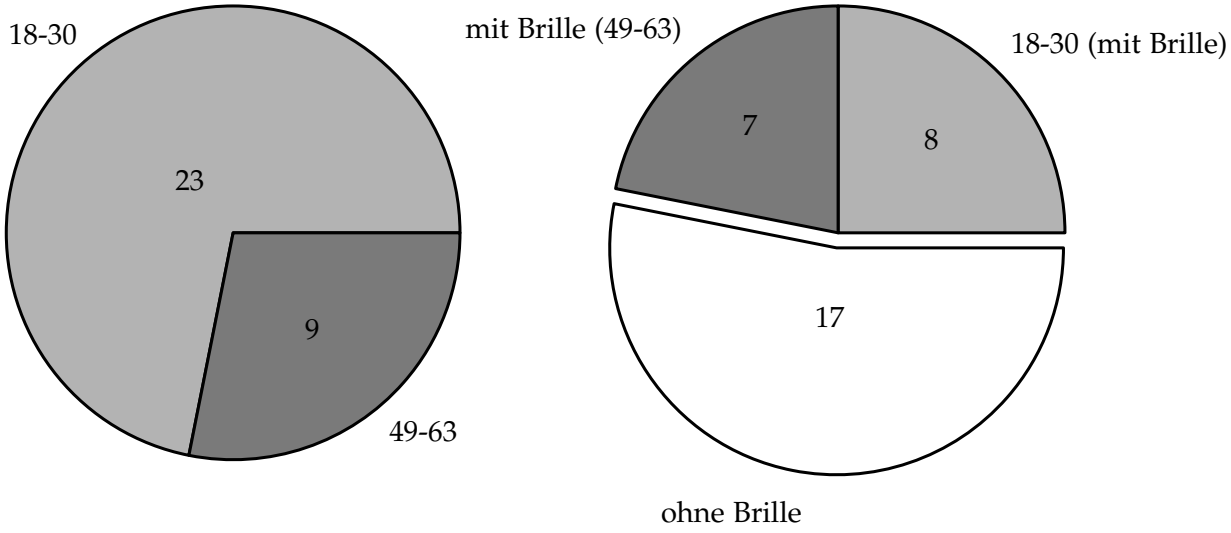


Abbildung 5.3.: Auswertung der Anzahl von Probanden in Jung und Alt, sowie Brillenträger

Probanden so gering ausfällt. Der Anteil ist etwa 70 % zu 30 % mit dem größeren Anteil bei jüngeren Probanden. In dem rechten Kreisdiagramm ist die Anzahl von Brillenträgern aufgezeichnet. Hier wird deutlich, dass etwas mehr als die Hälfte der Probanden keine Brille trug. Die Probanden, die eine Brille trugen, sind bei den beiden Gruppen etwa gleich aufgeteilt, jedoch aufgrund der voneinander abweichenden Gruppengrößen unterschiedlich innerhalb der Gruppen zu bewerten. Bei der Gruppe der älteren Probanden trugen etwa 80 % und bei der jüngeren Gruppe etwa 35 % eine Brille. Ein Proband trug Kontaktlinsen.

Auswertung des Versuchs aus psychologischer Sicht

Bei diesem Versuch wurden im Wesentlichen drei verschiedene Zeiten für die Probanden angeboten. Der Versuch startete entweder um 8 Uhr, 12 Uhr oder 15 Uhr. Um die verschiedenen Zeiten besser einzuordnen, kann in Abbildung 5.4 ein Histogramm über die Startpunkte der Versuche herangezogen werden. Hier wird deutlich, dass mit 14 Probanden gleich viele Versuche morgens zwischen 8 und 10 Uhr starteten wie auch mittags zwischen 11 und 14 Uhr. Am wenigsten Versuche haben nachmittags zwischen 15 und 16 Uhr stattgefunden, denn zu dieser Zeit haben nur vier Probanden teilgenommen. Außerdem variierte die Startzeit mittags etwas mehr, da 3 Versuche eine Stunde früher und zwei Versuche eine Stunde später stattfanden. Laut [Wee+98] gibt es zwei Zeiträume, in denen in anderen Versuchen

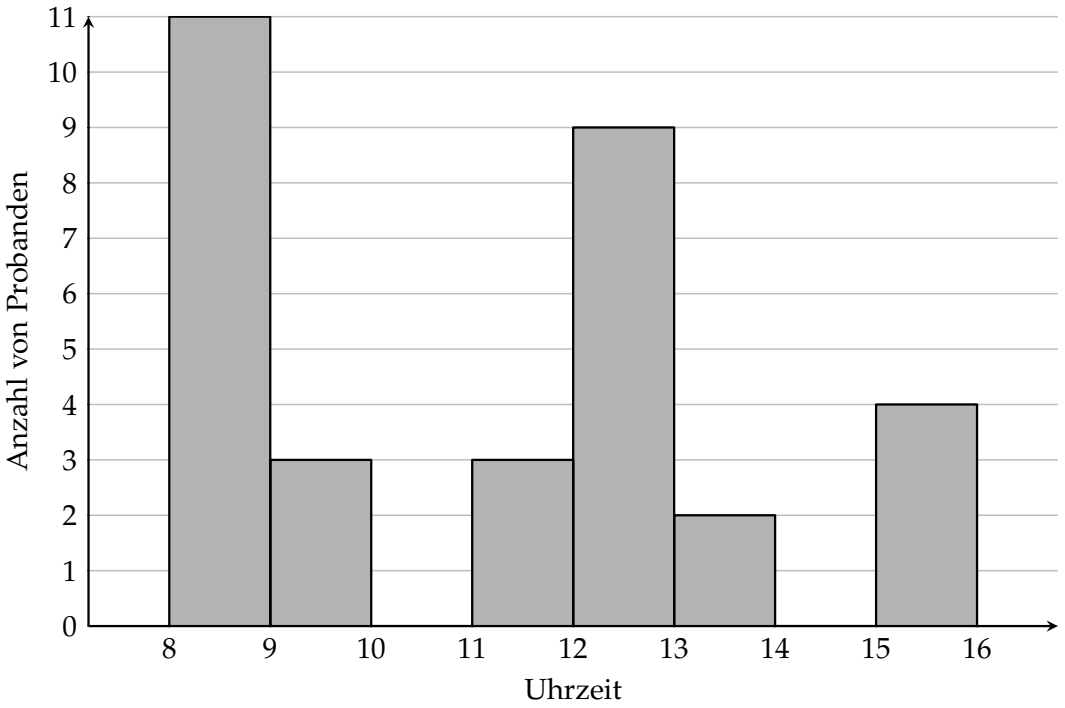


Abbildung 5.4.: Auswertung der Uhrzeiten an denen die Untersuchungen stattgefunden haben

eine Leistungsminderung festzustellen war: zwischen 3 und 4 Uhr morgens sowie zwischen 13 und 15 Uhr. In der hier vorliegenden Arbeit liegen zwei Versuche in diesem Zeitraum. Ebenso werden Zeiträume für erhöhte Leistungsbereitschaft beschrieben, die zwischen 7 und 11 Uhr und zwischen 16 und 20 Uhr liegen. In der hier vorliegenden Arbeit fallen 14 Probanden in den ersten genannten Bereich und könnten somit von erhöhtem Leistungsvermögen profitieren. Die Startzeitpunkte der restlichen 16 Probanden liegen in Zeiträumen, die weder für erhöhte noch verminderte Leistungen bekannt sind.

Bei der Einschätzung der Probanden nach der Fahrt, welche psychologischen Effekte subjektiv festzustellen waren, kam heraus, dass fast alle Probanden einen der gegebenen Zustände wahrgenommen haben. Die Anzahl der Probanden, die mindestens ein Mal während des Versuches einen der psychologischen Zustände wahrgenommen hat, wird in Abbildung 5.5 dargestellt. Der Zustand des Konzentrationsverlusts ist bei 94 % der Probanden eingetreten. Unaufmerksamkeit wurde von etwa 88 % der Probanden angegeben. Von Ablenkung waren nach eigenen Angaben 50 % betroffen. Am häufigsten wurde mit etwa 97 % Müdigkeit von den Probanden festgestellt. Da nur sehr wenig Literatur zu den letzten beiden Zuständen Sekundenschlaf und Tagträumen existiert, ist erstaunlich, dass doch viele Probanden diese Zustände bei der Selbsteinschätzung angaben. Dabei berichten etwa 44 % von Sekundenschlaf und etwa 34 % von mindestens einem Tagtraum. Ob diese Zustände tatsächlich eingetreten sind, lässt sich nur schwer beurteilen. Die genauen zeitlichen Einschätzungen können im Anhang A.2 eingesehen werden.

Wie in Kapitel 5.1 beschrieben, wird eine Befragung zur aktuellen Müdigkeit vor und

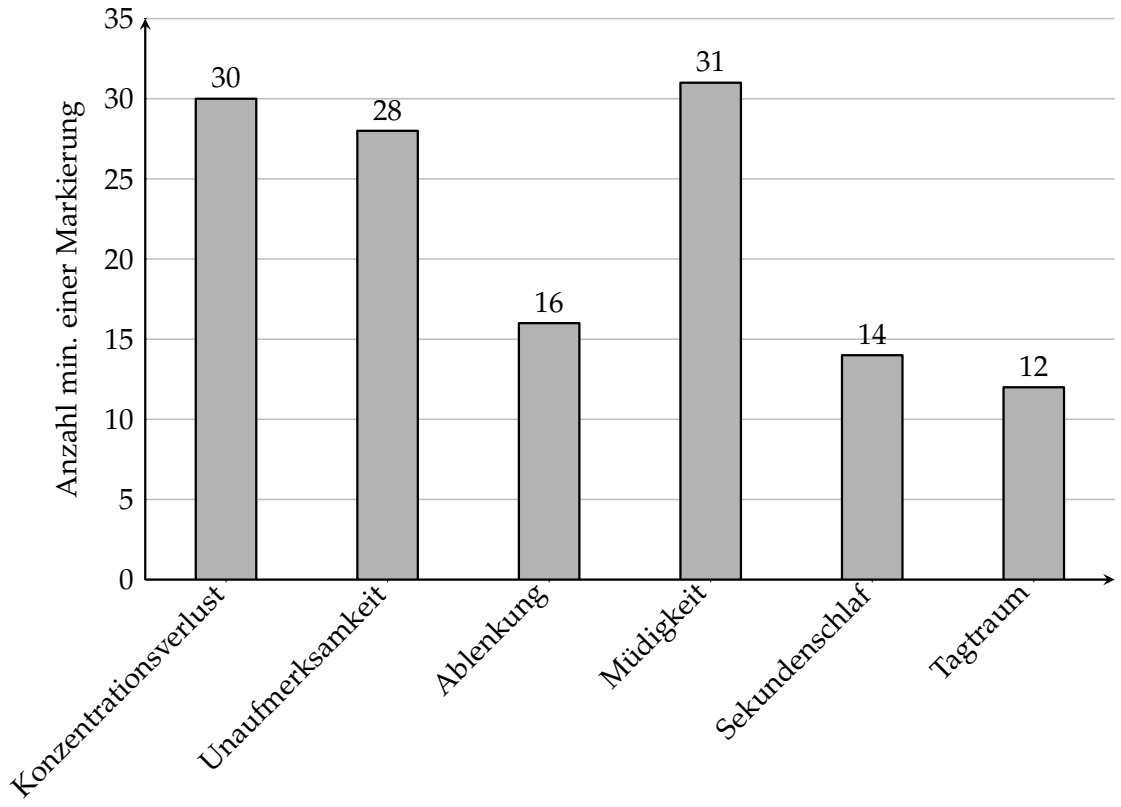


Abbildung 5.5.: Auswertung der Befragung zur Selbsteinschätzung des Zustands

nach der Fahrt durchgeführt. Hierzu wird eine statistische Auswertung dargestellt. Dabei handelt es sich um eine sogenannte Kastengrafik. Das Rechteck in dieser Darstellung weist auf die mittlere Verteilung von 50 % der Daten hin. Dazu wird in diesem Fall der Durchschnitt über alle Daten durch eine Raute und der Median durch eine grau gestrichelte Linie dargestellt. Die Antennen an den Rechtecken spiegeln außerhalb der Box liegende Daten wider. Mit einem x werden noch zusätzlich aufgetretene Ausreißer markiert. In Abbildung 5.6 wird diese Auswertung für den hier vorliegenden Versuch dargestellt. Als Gegenüberstellung werden links die Daten vor dem Versuch und rechts nach dem Versuch dargestellt. Hiermit wird deutlich, dass die Müdigkeit während des Versuchs wesentlich zunimmt. Der Durchschnittswert beträgt nach dem Versuch 2,75 Punkte mehr auf der neunstufigen Skala. Ebenso steigt der Median von 3 auf 7. Dazu nimmt die Streuung der Daten ab, was durch die kleinere Höhe des Rechtecks im Diagramm *nach der Fahrt* deutlich wird. Ein paar Ausreißer bei 4 und 3 nach der Fahrt sprechen dafür, dass diese wenigen Probanden vermutlich keine oder nur wenig Steigerung der Müdigkeit verspürt haben. Um für die Mittelwerte dieser Daten den Beweis für einen signifikanten Unterschied zu führen, müsste der von [Wel38] vorgestellte Signifikanztest (Welch-Test) angewendet werden. Der abgewandelte Zweistichproben-t-Test ist auf Stichproben unterschiedlicher Varianz anwendbar. Da dieser nur für Stichproben mit mehr als 30 Messpunkten geeignet ist, sind die Ergebnisse dieses Testes ungesichert. Somit ist für diese Daten leider keine Aussage über die Signifikanz zu treffen. Generell kann festgehalten werden, dass der Versuch bei den meisten Probanden Müdigkeit induziert. Jedoch

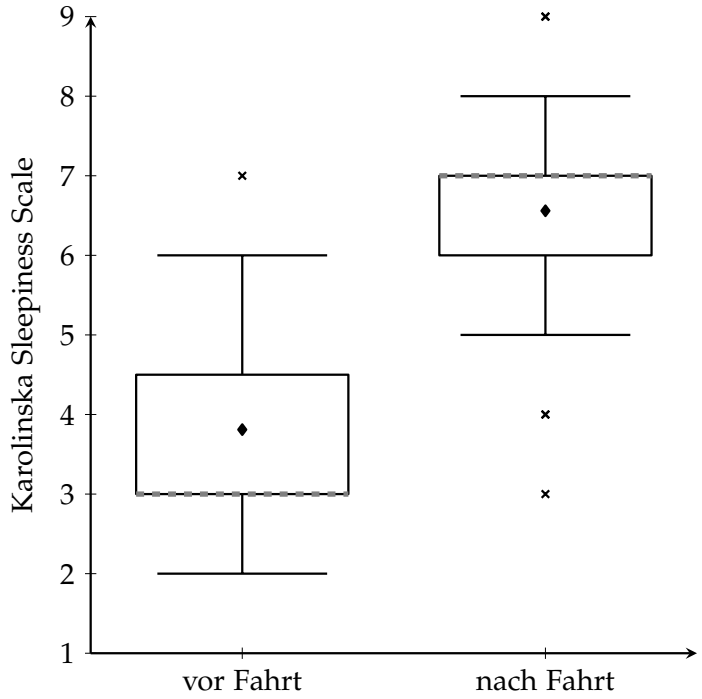


Abbildung 5.6.: Subjektive Einschätzung der Müdigkeit

liegt der Fokus der hier vorliegenden Arbeit auf Vigilanz, welche nicht mit Müdigkeit gleichzusetzen ist.

Um Vigilanz zu messen, ist die Definition aus [Mac48] dadurch gegeben, dass die Reaktionsbereitschaft vermindert wird. Dies hat sich in früheren Untersuchungen durch ausbleibende Reaktionen oder vergrößerte Reaktionszeiten verdeutlicht. Um die Vigilanz in der hier vorliegenden Arbeit zu untersuchen, wird eine Auswertung der Reaktionszeit wichtig, wobei sich in der Vorstudie gezeigt hat, dass die Reaktion auf einen visuellen Reiz schwer auszuwerten ist. Ebenso stellen die Hindernisse als Reize im Hauptversuch ein Problem dar. Um die Reaktionszeiten zu bestimmen, müsste jede Aufzeichnung aufwendig analysiert werden. Bei jedem Versuch ist der Verkehr etwas anders, sodass sich die Sichtweite von Proband zu Proband unterscheiden kann. Hier ist die Entscheidung gegen das Abspielen eines Videos gefallen, da bei der Vorstudie die Synchronisierung der Reaktionen des Probanden mit dem Video schwierig war. Die Reize der Spurwechsel sind ebenso problematisch. Hierbei gibt es Unterschiede bei jedem Probanden, da das Fahrzeug in einigen Fällen bereits auf der linken Spur fährt. Dazu gibt es Situationen, in denen das Fahrzeug nicht überholen kann, da die linke Spur belegt ist. Auch der genaue Zeitpunkt des Verstoßes ist nicht genau auszumachen, da nicht direkt aufgenommen werden kann, wann das Fahrzeug die Linie überfährt. Dahingegen ist die Messung von Reaktionszeiten bei Geschwindigkeitsverstößen recht einfach zu gestalten. Die Sichtbarkeit eines Schildes stellt keinen Grund für eine Reaktion dar, denn die Regelung gilt erst bei Passieren des Schildes. Diese Aktion kann in der Fahrsimulation mit einem Zeitstempel festgehalten werden. Ebenso wird die Reaktion des Probanden mit einem Zeitstempel versehen, sodass eine Berechnung der Reaktionszeit direkt

möglich ist. Bei der automatischen Auswertung der Reaktionszeiten sind, ähnlich wie in der Vorstudie, negative Reaktionszeiten aufgetreten, welche in diesem Fall jedoch ausgeschlossen werden können. Bei dem Hauptversuch ist es nicht möglich, vor dem Schild zu reagieren, da sich der Reiz erst nach Passieren des Schildes einstellt. Dazu sei darauf hingewiesen, dass das Fahrzeug erst eine kurze Zeit (wenige Sekunden) nach Passieren des Schildes das Verhalten auch in normalen Situationen anpasst, sodass die Reaktionszeiten dementsprechend größer ausfallen können. Die in diesem Versuch ermittelten Reaktionszeiten für den Reiz der Überschreitung der Geschwindigkeitsbegrenzung kann in Abbildung 5.7 gesehen werden. An dieser Stelle werden wieder Kastengrafiken dargestellt, die ebenfalls, wie oben beschrieben, gestaltet sind. Auf der x-Achse sind die drei Reize mit der Nummer und auf der y-Achse die Reaktionszeit in Sekunden aufgetragen. Um zunächst auf den Median (grau gestrichelt) einzugehen, welcher weniger von Ausreißern beeinflusst wird als der Mittelwert, kann zwischen Reiz 1 und 2 ein leichter Abfall festgestellt werden. Hier besteht eine Ähnlichkeit zu den Ergebnissen aus der Vorstudie. Dies könnte, wie zuvor schon erwähnt, durch den Trainingseffekt ausgelöst werden. Gleichzeitig wird hier allerdings deutlich, dass die Streuung der Reaktionszeiten zugenommen hat, was durch die größere Höhe des Rechtecks bei Reiz 2 im Gegensatz zu Reiz 1 sichtbar wird. Hier könnte ein Zeitpunkt eingetreten sein, bei dem für einige Probanden bereits die Vigilanz deutlich wird, jedoch andere noch nicht davon beeinflusst werden. Wird nun Reiz 3 herangezogen, wird deutlich, dass der Median der Reaktionszeiten wesentlich ansteigt. Dieser liegt jetzt etwa auf der Höhe der Antenne von Reiz 1 und über der oberen 50 %-Grenze des Kastens von Reiz 2. Bei Reiz 3 ist der Kasten, der die Hälfte der Reaktionszeiten widerspiegelt wieder etwas weniger hoch als bei Reiz 2, was für eine geringere Varianz in den Daten spricht. Für die Reize 1 und 2 liegen die mit einer Raute markierten Mittelwerte recht nahe beieinander, jedoch für Reiz 2 etwas geringer. Bei Reiz 3 wird deutlich, dass der Mittelwert aufgrund der Ausreißer die Daten nicht so gut repräsentieren kann. Dieser liegt deutlich über der oberen Kante des Kastens. Dazu muss darauf hingewiesen werden, dass in dieser Darstellung ein Ausreißer bei Reiz 3 nicht dargestellt wird. Dieser liegt bei 28,22 s und würde somit die Darstellung sehr viel unübersichtlicher machen.

Neben den größer werdenden Reaktionszeiten sind ausbleibende Reaktionen auf Reize ebenso ein relevantes Maß für die Vigilanz. Die maximale Zeit, in der ein Proband ordentlich auf einen Reiz reagieren kann, ist auf 30 s begrenzt, da das Fahrzeug nach dieser Zeit ein weiteres Schild passiert. Gibt es in dieser Zeit keine Reaktion von dem Probanden, wird dies als ausbleibende Reaktion gewertet. Hierzu kann in Abbildung 5.8 die Auswertung für die Hauptstudie gesehen werden. Dabei sind auf der x-Achse wieder die 3 Reize dargestellt, die y-Achse zeigt die prozentuale Anzahl ausbleibender Reaktionen und in grauen Säulen ist die Auswertung über alle Probanden gezeigt. Die schwarzen Säulen zeigen die junge sowie die gepunkteten Säulen die ältere Gruppe. Wird nun die Auswertung über alle Probanden betrachtet, kann eine Zunahme der prozentualen Anzahl ausbleibender Reaktionen festgestellt werden. Bei Reiz 2 ist die prozentuale Anzahl ausbleibender Reaktionen neun Prozentpunkte höher als bei Reiz 1. Diese Steigerung setzt sich fort, denn bei Reiz 3 ist der prozentuale Ausfall um weitere 22 Prozentpunkte höher. Bei diesem Reiz sind 44 Prozent der Probanden nicht

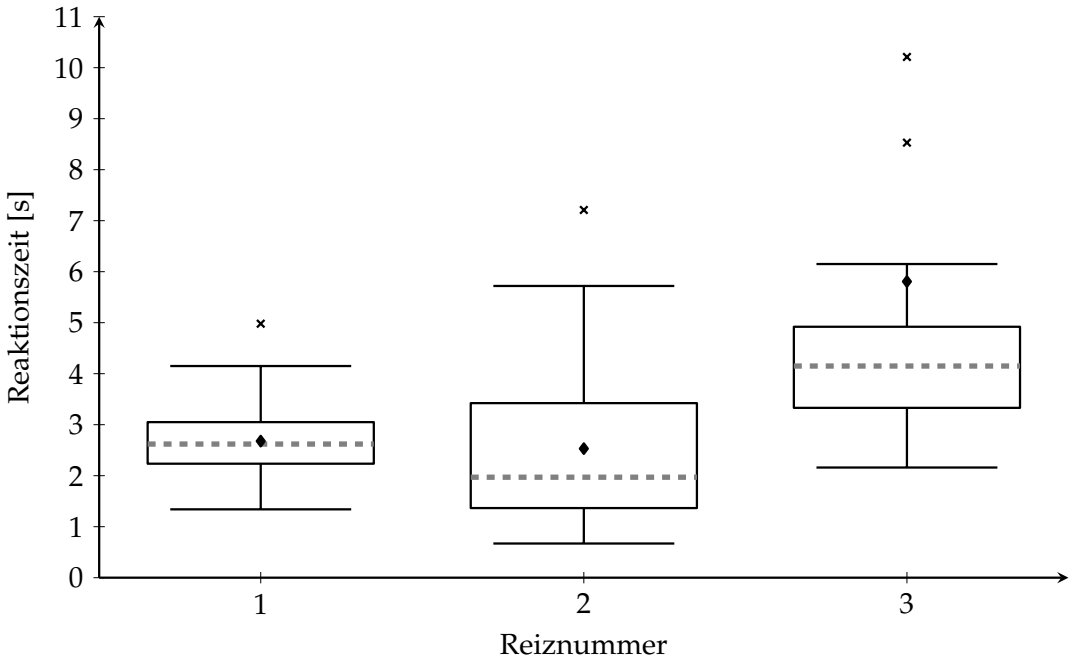


Abbildung 5.7.: Reaktionszeiten der Probanden auf Missachten der Geschwindigkeitsbegrenzung

in der Lage, auf den gegebenen Reiz zu reagieren. Hiermit wird also, ebenso wie bei den Reaktionszeiten, deutlich, dass in dem Versuch Vigilanz ausgelöst wird. Werden die Zahlen für die jüngere Gruppe miteinander verglichen, wird hier ein ähnliches Ergebnis deutlich. Die prozentuale Anzahl ausbleibender Reaktionen steigt von Reiz 1 zu Reiz 2 von 9 % um acht Prozentpunkte auf 17 % an. Von Reiz 2 auf Reiz 3 wird dieser Effekt noch deutlicher, denn hier steigt die prozentuale Anzahl ausbleibender Reaktionen um 31 Prozentpunkte auf 48 % an. Das bedeutet, dass nahezu die Hälfte der jüngeren Probanden nicht auf den Reiz in Minute 56 reagiert hat. Bei der älteren Gruppe ist nicht so eine große Steigerung zu sehen, dabei ist die Gruppe jedoch auch wesentlich kleiner, sodass einzelne Probanden einen großen Sprung in der relativen Anzahl bedeuten. Hier gibt es zwischen Reiz 1 und 2 eine kleine Steigerung, denn ein Proband weniger konnte auf den Reiz regieren, sodass die prozentuale Anzahl ausbleibender Reaktionen von 22 % um elf Prozentpunkte auf 33 % ansteigt. Zwischen dem Reiz 2 und 3 gibt es dann bei dieser Gruppe keine Veränderung mehr. Die prozentuale Anzahl ausbleibender Reaktionen bei dem ersten Reiz könnte zudem darauf hinweisen, dass die Aufgabe nicht richtig verstanden wurde oder ein Fehler bei der Aufzeichnung passierte. Reiz 1 findet innerhalb der ersten fünf Minuten statt, sodass davon ausgegangen werden kann, dass hier die Probanden noch vigilant sein sollten.

Zusammenfassend kann gesagt werden, dass die psychologischen Messungen ergeben, dass in dieser Studie Vigilanz aufgetreten ist. Zum einen steigen die Reaktionszeiten an, zum anderen nimmt die Anzahl der ausbleibenden Reaktionen zu. Beides spricht für den Effekt der Vigilanz, wie er in der Literatur beschrieben wird. Dabei ist dieser Effekt besonders bei der jüngeren Gruppe aufgetreten, wobei die ältere Gruppe zu

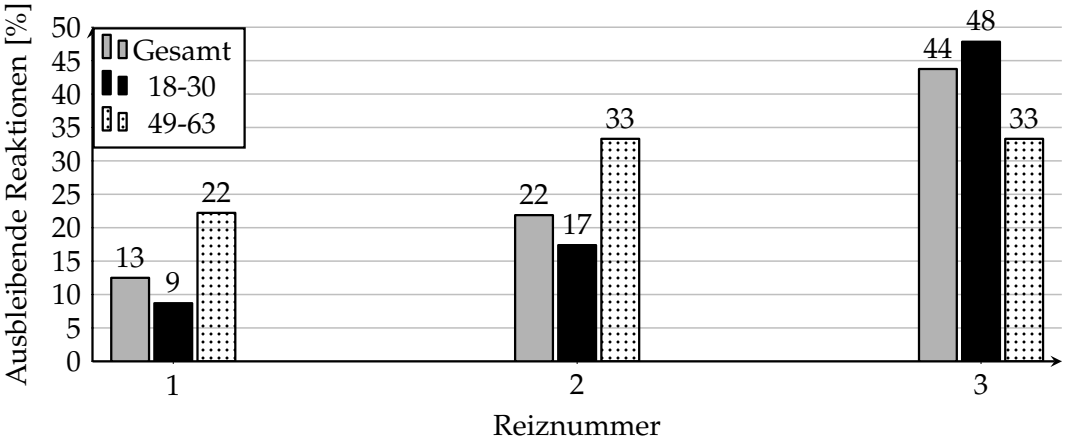


Abbildung 5.8.: Ausbleibende Reaktion bei der Überschreitung der Geschwindigkeitsbegrenzung

klein ist, um auf eine allgemeine Repräsentanz zurückschließen zu können. Somit kann davon ausgegangen werden, dass das Überwachen der Fahraufgabe zu einem Problem werden könnte, da ein Fahrer nach längerer Zeit nur noch vermindert auf solche wichtigen Reize reagieren kann. Die Reize haben sich hierbei jedoch leicht unterschieden, was zusätzlich Einfluss auf die Reaktionen nehmen kann. Dazu muss beachtet werden, dass der Versuch in einem statischen Fahrsimulator stattgefunden hat. Die Simulation wirkt sich auf das Vertrauen in die Automatisierung sowie die körperliche Wahrnehmung der Reize aus. Ein haptisches Feedback könnte dazu führen, dass die Probanden einen Geschwindigkeitswechsel schneller bemerken als in dieser Simulation.

6

Datenverarbeitung und Fusion von verschiedenen physiologischen Merkmalen

In diesem Kapitel soll es um die Datenverarbeitung aus elektrotechnischer Sicht gehen. Hierbei werden die Erkenntnisse aus der psychologischen Analyse dazu genutzt, um die Daten einordnen zu können. Ziel dabei ist eine automatisierte Datenverarbeitung mithilfe eines Modells so zu erstellen, dass der psychologische Zustand später maschinell erfasst werden kann. Hierbei werden, in der vorliegenden Arbeit, zwei verschiedene Ansätze vorgestellt und im nächsten Kapitel ausgewertet und miteinander verglichen. Zum einen wird die Fusion der verschiedenen Sensoren auf Merkmalsebene durchgeführt und zum anderen werden die Daten erst nach der Klassifizierung auf der Entscheidungsebene fusioniert. Zunächst wird das Vorgehen mit den Daten aus der Vorstudie vorgestellt und im Anschluss werden die Daten aus der Hauptstudie für eine weitere Auswertung herangezogen.

6.1. Datenverarbeitung zur Erstellung von Merkmalen aus physiologischen Daten

Dieser Abschnitt legt den Fokus darauf, wie die Daten, die von den Sensoren aufgenommen wurden, weiter verarbeitet werden. Dabei geht es zunächst um die Vorbereitung der Daten, indem diese zeitlich synchronisiert und dann teilweise statistisch ausgewertet werden. Es werden sowohl die Daten aus der Vorstudie als auch aus dem Hauptversuch behandelt. Der größte Unterschied zwischen den beiden Datensätzen besteht darin, wie die Einordnung der Daten in die verschiedenen Klassen erfolgt. Zuletzt wird für jeden Datensatz noch beschrieben, wie die Daten in die verschiedenen Gruppen Training, Test und Validierung aufgeteilt werden.

Datenverarbeitung und Einordnung der Daten aus der Vorstudie

Da die Sensoren unterschiedliche Abtastraten aufweisen, muss der Datensatz auf eine gemeinsame Zeitskala mit definierten Abständen gebracht werden. Dabei wird

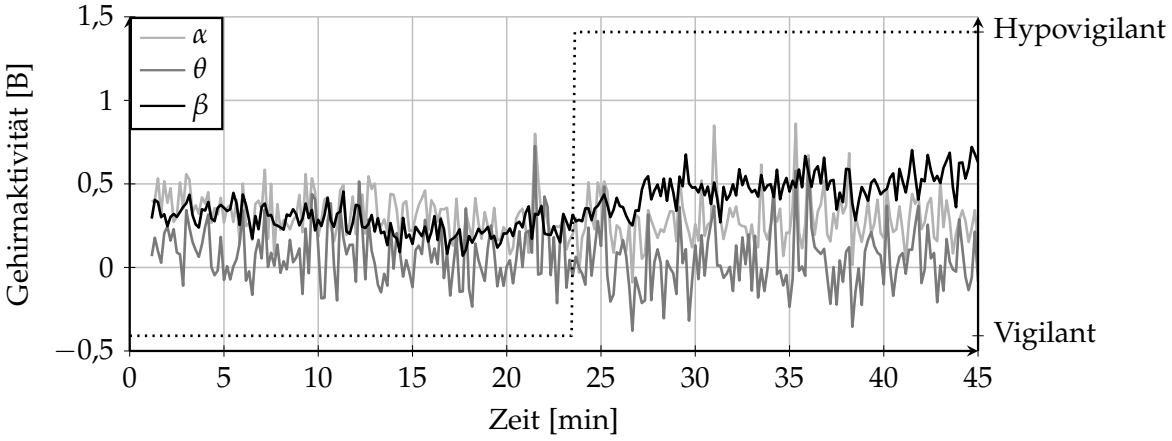


Abbildung 6.1.: Gehirnaktivität α, θ und β eines Probanden aus dem Vorversuch mit eingezeichneter Beschriftung (Ground Truth)

eine Abtastzeit von zehn Sekunden gewählt, da die Mittlung der Gehirnaktivität die kleinste Abtastzeit aller Messgrößen darstellt. Die anderen Messgrößen werden in einem Intervall von 60 s bestimmt.
Um die Messwerte für diese festen Zeitschritte zu bestimmen, wird eine lineare Interpolation der Messdaten zwischen benachbarten Zeitstempeln durchgeführt. Das bedeutet, dass die Messung eines Probanden über 45 Minuten (2700 s) 270 Datenpunkte erzeugt. Da in der ersten Minute noch starke Bewegungsartefakte das EEG stören, werden diese Messwerte gelöscht, sodass 264 Datenpunkte weiter verarbeitet werden können. Für das später verwendete überwachte Lernverfahren ist es notwendig die Daten in die verschiedenen Klassen, hier Vigilant und Hypovigilant genannt, einzuordnen. Wie bereits zuvor beschrieben, gibt es keinen psychologischen Beweis in der Vorstudie für diese Zustände. Um die Daten dennoch verwenden zu können, wird die Gehirnaktivität herangezogen, um die Daten in diese beiden Zustände einzuordnen.
Dabei wird bei einer großen Veränderung der β- und γ-Wellen auf einen Zustandswechsel geschlossen.
So kann in Abbildung 6.1 und 6.2 das Vorgehen an einem Beispiel anhand der Trainingsdaten gesehen werden. Die Gehirnaktivität weist ab etwa Minute 24 eine größere Veränderung auf, da die β- und γ-Wellen im Vergleich zu den anderen Gehirnaktivitäten ansteigen. Dabei wird darauf geachtet, dass die Mittelwerte sich deutlich voneinander unterscheiden (Mittelwertänderung größer als Varianz im Signal). So wird in diesem Fall auch die Kennzeichnung des kognitiven Fahrerzustands als gepunktete Linie dargestellt. Ab diesem Zeitpunkt wird die Beschriftung von dem hier sogenannten Zustand Vigilant zu Hypovigilant geändert. Diese Annahme fußt auf der oben erwähnten subvigilen β-Aktivität, welche mit einer Verminderung der Vigilanz nach [Zsc02] einhergeht. Dazu ist in [Haa07] zusammengefasst, dass eine γ-Aktivität mit einem Wachzustand verbunden werden kann, sodass hier möglicherweise keine Müdigkeit, sondern eine angestrengte Aufmerksamkeit mit Wachheit verbunden sein könnte. Leider gibt es aber zu diesem Effekt keine weiteren Untersuchungen in der Literatur, sodass hier ein recht

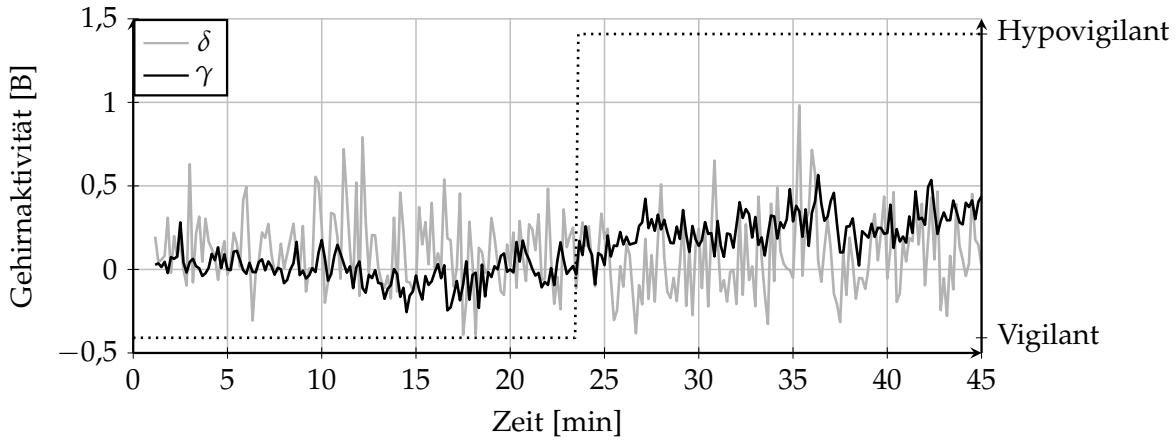


Abbildung 6.2.: Gehirnaktivität δ und γ eines Probanden aus dem Vorversuch mit eingezeichneter Beschriftung (Ground Truth)

explorativer Ansatz gewählt wird, der im Nachhinein noch von weiteren Studien belegt werden sollte. Die weiteren Daten werden nach dem gleichen Prinzip manuell beschriftet, was aufgrund der individuellen Unterschiede nicht immer präzise durchgeführt werden kann, wie es für β-Aktivitäten schon von [Zsc02] beschrieben wurde. So kann bei einigen Probanden der Zeitpunkt des Zustandswechsels zeitlich nur grob bestimmt werden, da die Änderung der Gehirnaktivität bei jedem Probanden unterschiedlich stark ausgeprägt ist. Es gibt durch die deutlichen Unterschiede zwischen den einzelnen Probanden keine festen Schwellenwerte, jedoch sind qualitative Aussagen der Probanden nach dem Versuch mit in die Bewertung eingeflossen. Eine andere Möglichkeit, Unterschiede auszumachen, wäre die Analyse einer signifikanten Änderung in den Sensordaten, jedoch sind hierfür nicht genügend Datenpunkte pro Proband vorhanden.
Nach dem Erkennen von starken Änderungen und dem Beschriften der Daten muss noch dafür gesorgt werden, dass die Werte jedes Sensors in gleicher Weise von dem Netzwerk verarbeitet werden können. Dazu werden alle Daten von dem Mittelwert befreit und über die Varianz normiert. So wird sichergestellt, dass kein Merkmal vom Netzwerk bevorzugt wird und die Gewichte sich in der gleichen Größenordnung befinden. Die Daten der neun Probanden werden in Trainings- und Testdaten aufgeteilt. Dabei werden die Daten von acht Probanden für die Trainingsdaten verwendet und die Daten eines Probanden als Testdaten herangezogen. So werden etwa 10 % der Daten als Testdaten reserviert. Da die Gesamtgröße der Gruppe so klein ist, wird in diesem Fall auf das Abgrenzen von Validierungsdaten verzichtet. In den folgenden Abschnitten bleibt diese Aufteilung gleich.

Datenverarbeitung und Einordnung der Daten aus dem Hauptversuch

Die zeitliche Synchronisation funktioniert in diesem Fall analog zu der Verarbeitung der Daten aus dem Vorversuch. Auch hier werden die Daten mit den entsprechenden Zeitstempeln zusammengebracht und dann weiter verwendet. Um die Ergebnisse vergleichbar zu der Vorstudie zu halten, wird in diesem Fall ebenfalls eine

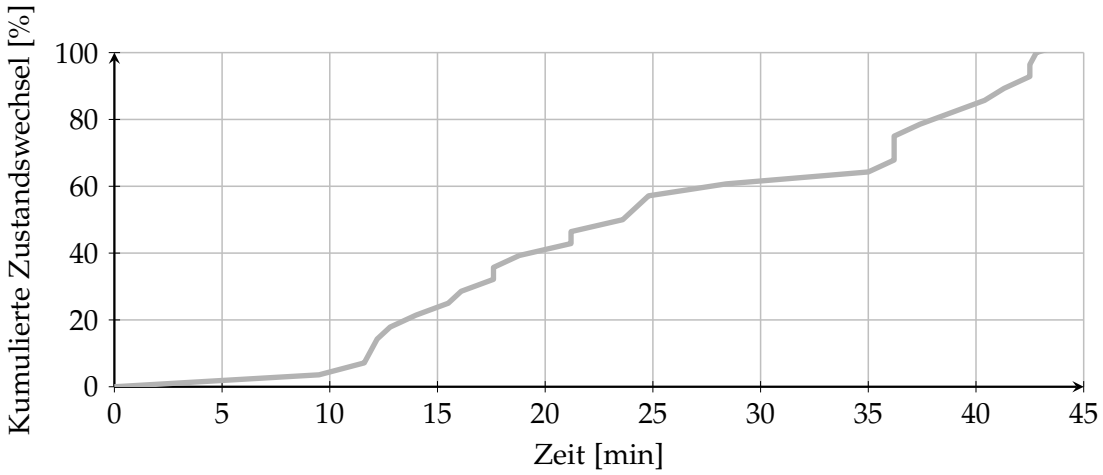


Abbildung 6.3.: Prozentuale kumulierte Summe der Zustandswechsel ($N = 30$)

Interpolation auf ein Intervall von zehn Sekunden vorgenommen. In diesem Fall wäre eine Abtastung von minimal einer Sekunde möglich, da dies die kleinste Messzeit des Herzens ist. Eine höhere Abtastung würde jedoch auch eine größere Herausforderung für das Modell mit sich bringen. Dies ist ein weiteres Argument dafür, eine größere Abtastzeit zu wählen. So entstehen bei einer Messung eines Probanden über 60 Minuten (3600 s) 360 Datenpunkte. Da die Messung der Sensoren vor Beginn der Simulation gestartet wird, ist es nicht nötigt, die 360 Datenpunkte am Anfang zu reduzieren, wie in der Vorstudie der Fall war.

Auch die Daten aus dem Hauptversuch müssen für das überwachte Lernverfahren eingeordnet werden. Dabei wird die in Kapitel 5.2 gezeigte Auswertung der subjektiven Befragung der Probanden genutzt. Am Ende hat jeder Proband sich selbst zeitlich eingeschätzt und angegeben, zu welchem Zeitpunkt die Konzentration verloren gegangen ist. Diese Information wird nun verwendet, um den Zeitpunkt des Zustandswechsels zu bestimmen.

In Abbildung 6.3 sind die Zeitpunkte der Zustandswechsel abgebildet. Hier ist auf der x-Achse die Versuchszeit aufgetragen und auf der y-Achse ist die relative kumulierte Summe der Zustandswechsel von 30 Probanden aufgetragen. Wie im Anhang in der Abbildung A.4 genauer zu sehen ist, haben zwei Probanden keine Selbsteinschätzung abgegeben und werden somit nicht weiter betrachtet. Deutlich wird, dass schon nach 23,6 Minuten die Hälfte der Probanden einen Konzentrationsverlust eingeschätzt hat. Eine große Steigerung in dieser Abbildung kann zwischen Minute 35 und 45 gesehen werden, wo etwa 40 % der Probanden einen Zustandswechsel angeben. Auch hier wird nach dem Beschriften dafür gesorgt, dass die Daten für die weitere Verarbeitung im Netzwerk normiert werden. Hierzu werden diese ebenfalls vom Mittelwert befreit und über die Varianz normiert. Damit sollten sich auch hier die Gewichte im Netzwerk in gleichen Größenordnungen bewegen. Um die Balance zwischen den beiden Klassen nicht noch weiter zu verschlechtern (wie aus Abbildung 6.3 hervorgeht), werden die zwei Probanden ohne Selbsteinschätzung aus den Trainings- und Testdaten ausgeschlossen. Ebenso gab es bei einigen Probanden Probleme bei der Erfassung der Körperfunktionen. Dabei fehlen die Daten ab einem bestimmten

Zeitpunkt komplett oder sind nur teilweise verloren gegangen. Ein Auslöser hierfür war in einem Fall, dass sich der Herzsensor gelöst hatte. Somit werden bei den folgenden Untersuchungen die Probanden mit den Identifikationsnummern (ID) 7, 10, 14, 17 und 20 nicht weiter berücksichtigt. Da genügend Probanden zur Verfügung standen, können in diesem Fall die Daten von drei Probanden als Testdaten verwendet werden. Dabei werden die Probanden mit der ID 1-3 ausgewählt. Zusätzlich werden für eine Auswertung des Trainings noch weitere vier Probanden separiert, die als Validierungsdaten fungieren sollen. So ergibt sich für das Training ein Datensatz mit 18 Probanden, der im Folgenden bei jeder Untersuchung zur Hauptstudie verwendet werden wird. In diesem Fall werden also 12 % der Daten als Testdaten, 16 % der Daten als Validierungsdaten und die restlichen 72 % als Trainingsdaten ausgewiesen.

6.2. Fusion der physiologischen Daten auf Merkmalsebene

In diesem Abschnitt soll die Struktur der Modelle erläutert werden, die eine Fusion auf Merkmalsebene durchführen. Dabei wird sowohl auf das Modell für den Datensatz aus dem Vorversuch als auch auf das Modell für den Datensatz aus dem Hauptversuch eingegangen. Diese unterscheiden sich sowohl in der Anzahl der Probanden als auch in den Merkmalen, die für die Analyse verwendet werden können. Da bisher kein analytischer Zusammenhang zwischen den physiologischen Körperfunktionen und den hier untersuchten psychologischen Zuständen existiert, werden in der hier vorliegenden Arbeit künstliche neuronale Netzwerke verwendet. Da davon ausgegangen werden kann, dass sich Müdigkeit kumuliert, ist es wichtig, den zeitlichen Verlauf der Merkmale zu berücksichtigen, sodass bei der Struktur die Wahl auf Long Short-Term Memory (LSTM) Netzwerke [HS97] gefallen ist. Diese wurden zum Beispiel auch schon von [Xia+19] zur Erkennung von Müdigkeit, ähnlich wie in der vorliegenden Arbeit, verwendet.

Modell mit Gedächtnis auf Basis vom Datensatz der Vorstudie

In Abbildung 6.4 ist die Struktur des vollständigen Netzwerks für die Vorstudie abgebildet, die aus einer LSTM-Schicht, einer vollständig verknüpften Schicht (FC) und einer SoftMax-Schicht besteht. Der Eingang wird mit Sequenzen mit einer Länge von 264 Messpunkten, wie in Abschnitt 6.1 beschrieben, belegt. Die Anzahl der Sequenzen bestimmt sich aus der Anzahl der Merkmale, die für die Klassifizierung benutzt werden. In diesem Fall ist das Netzwerk mit den drei Merkmalen aus der Tabelle 6.1 dargestellt, die sich in zwei Herzmerkmale und einem Augenmerkmal aufteilen. Die LSTM-Schicht besteht aus 300 versteckten Einheiten, die parallel mit den Sequenzen gespeist werden. Es entstehen 300 Sequenzen, die durch die LSTM-Schicht verarbeitet wurden, die nun als Eingang für das vollständig verknüpfte Netzwerk dienen. Dieses fungiert als Klassifikator, sodass es eine Sequenz von Tupeln ausgibt, welche durch die SoftMax-Schicht zu einer Sequenz von Klassifikationsergebnissen verarbeitet wird. Somit wird mit diesem Ansatz ein Netzwerk erstellt, welches die Informationen aus allen Sensoren benötigt, um eine Klassifikation des Fahrerzustands abzugeben.

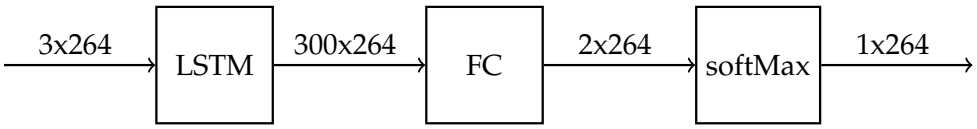


Abbildung 6.4.: Struktur des künstlichen neuronalen Netzwerks

Tabelle 6.1.: Merkmale für das Training der Datenverarbeitung

	Körperteil	Merkmal
1	Herz	Herzfrequenz HF
2		Herzraten-Variabilität H
3	Augen	PERCLOS

Das Training wird über eine Anzahl von 150 Epochen durchgeführt. Diese Anzahl wird so groß gewählt, dass der Fehler nur noch einen kleinen Gradienten besitzt und sich kaum mehr verbessert, wie in Abbildung 6.5 zu sehen ist.
Die initiale Lernrate wird auf 0,02 festgesetzt, welche nach der Hälfte der Epochen noch einmal auf 0,002 verkleinert wird. Dies spiegelt sich im Trainingsfehler in Abbildung 6.5 wider. Hier ist zu erkennen, dass die Varianz des Fehlersignals deutlich abnimmt. Eine Auswertung zur Validierung des Trainings ist in diesem Fall nicht möglich, da die Datenmenge zu gering ist, um hierfür Daten bereitzustellen.

Modell mit Gedächtnis auf Basis vom Datensatz der Hauptstudie

In Abbildung 6.11 ist die Struktur des vollständigen Netzwerks für den Datensatz der Hauptstudie zu sehen. Hierbei gibt es Unterschiede sowohl in der Struktur als auch in der Anzahl der Merkmale. Bei der Struktur ist in diesem Fall noch eine Dropout-Schicht hinzugefügt. Diese wurde zwischen der LSTM-Schicht und der vollständig verknüpften Schicht eingefügt. Die Softmax-Schicht bleibt wie bei der Vorstudie bestehen. In diesem Fall ändert sich aufgrund der längeren Messzeit auch die Länge der Sequenzen, welche in diesem Fall aus 360 Messpunkten besteht. Die zweite Dimension der Sequenzen wird durch die Anzahl der Merkmale bestimmt, die in Tabelle 6.2 aufgelistet ist. In diesem Fall kommen zu den drei bestehenden Merkmalen aus der Vorstudie nun 13 weitere Merkmale hinzu. Zwölf von den neuen Merkmalen stammen von den Sitzdrucksensoren. Zu den Messpositionen der Sitzdrucksensoren kann die Abbildung A.3 herangezogen werden. Ein Merkmal kommt bei der Messung der Augen, mit der Blinzel-Häufigkeit, hinzu. So ist die Gesamtzahl der Merkmale in diesem Fall auf 16 erhöht. Beispiele für die Werte, der einzelnen Merkmale über die Zeit von einem Probanden aus den Testdaten sind in den Abbildungen 6.6 bis 6.10 zu sehen. In Abbildung 6.6 ist die Herzraten-Variabilität dargestellt, welche im Mittel keine sichtbare Steigung annimmt. Auch die Varianz dieses Signals scheint über die komplette Zeit recht ausgeglichen. Das Merkmal Herzfrequenz HF ist in Abbildung 6.7 zu sehen. Bei diesem Signal kann im Mittel ein

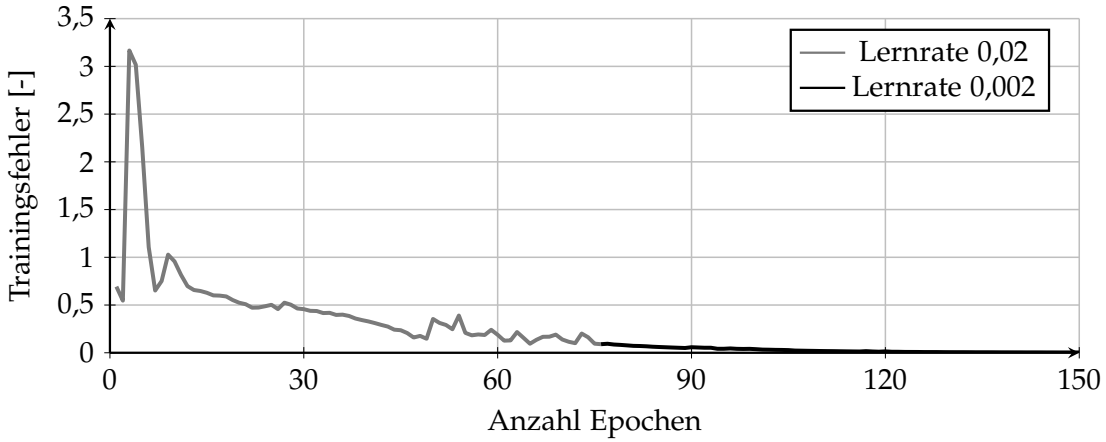


Abbildung 6.5.: Wert der Fehlerfunktion beim Training mit drei Merkmalen auf den Trainingsdaten von acht Probanden über die Anzahl der trainierten Epochen

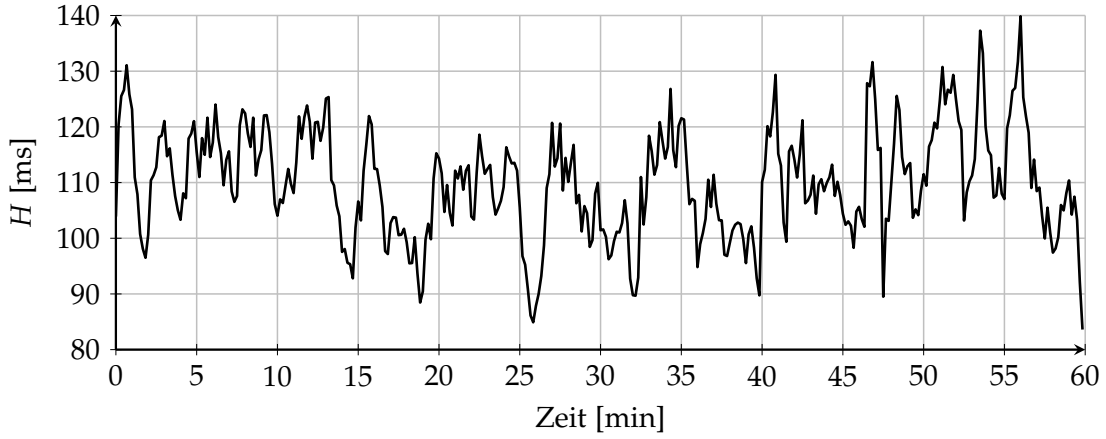


Abbildung 6.6.: Herzraten-Variabilität H des Probanden mit der ID 1 aus der Vigilanzstudie

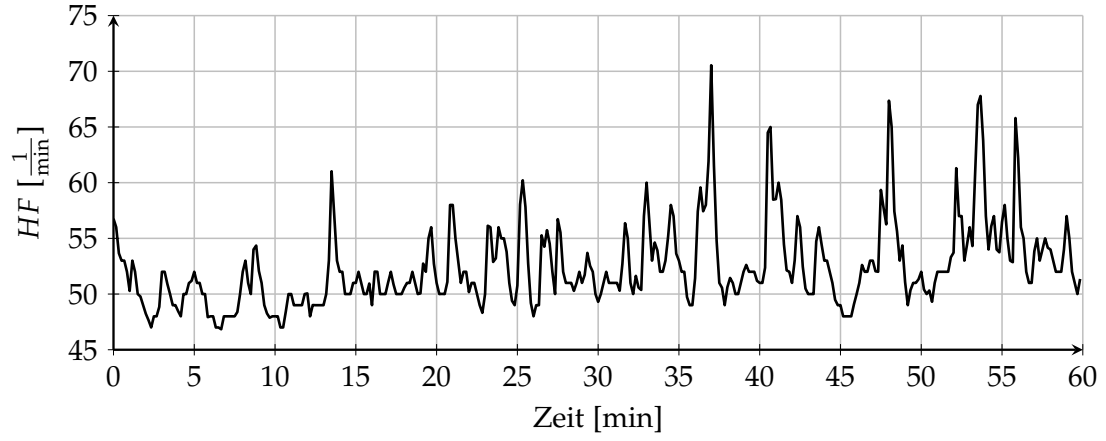


Abbildung 6.7.: Herzfrequenz HF des Probanden mit der ID 1 aus der Vigilanzstudie

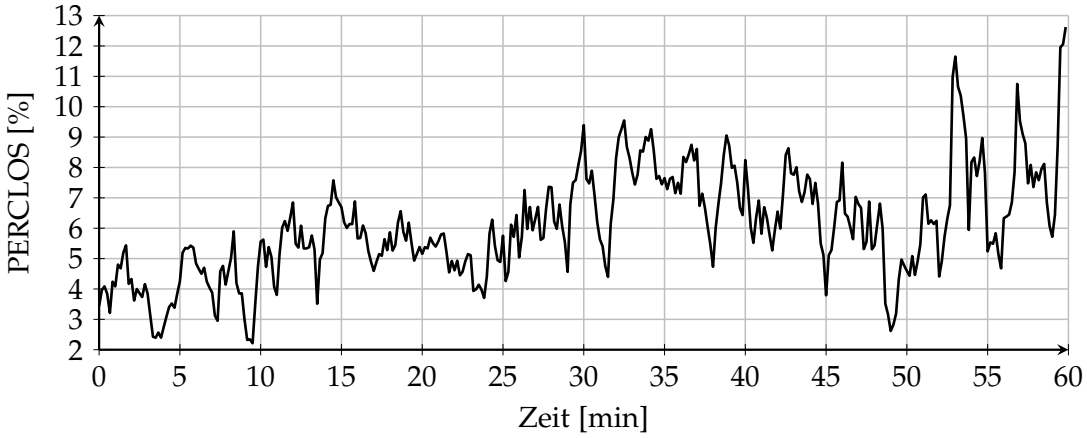


Abbildung 6.8.: PERCLOS Wert des Probanden mit der ID 1 aus der Vigilanzstudie

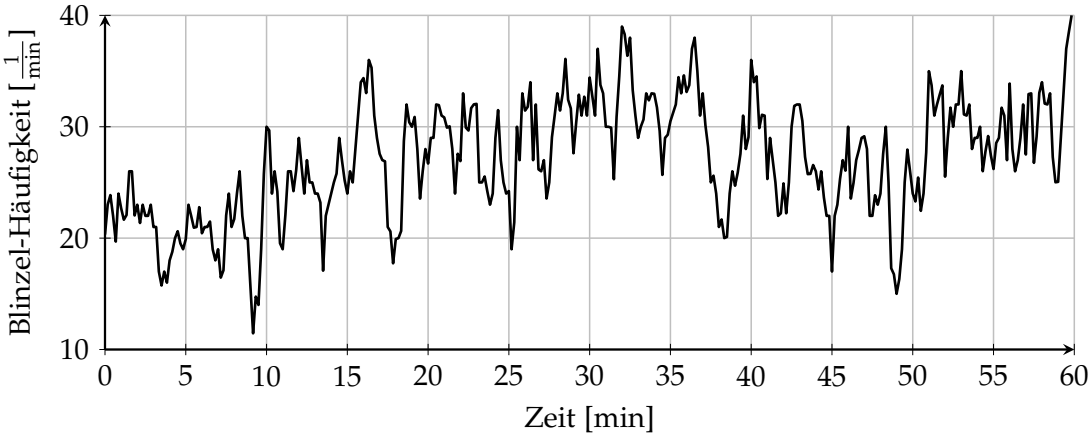


Abbildung 6.9.: Blinzel-Häufigkeit des Probanden mit der ID 1 aus der Vigilanzstudie

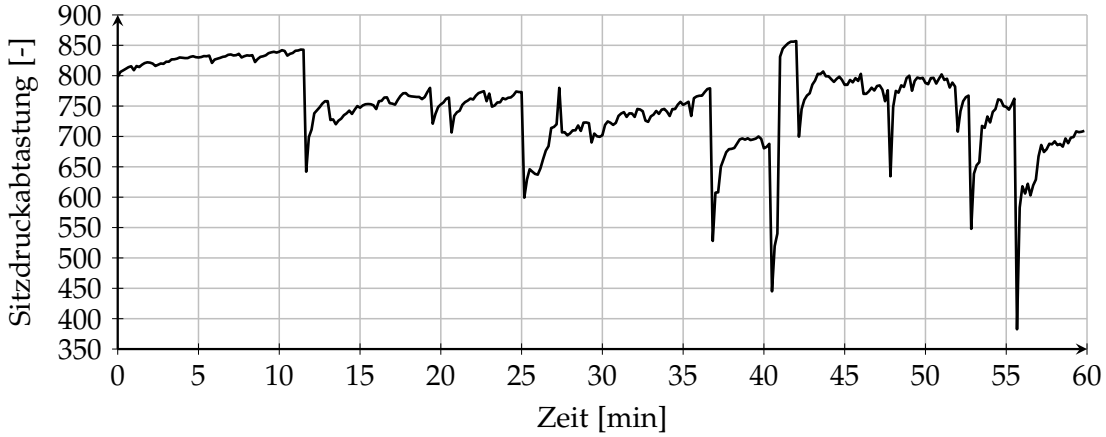


Abbildung 6.10.: Daten eines Sitzdrucksensors (mit der Nummer 6) des Probanden mit der ID 1 aus der Vigilanzstudie

leichter Anstieg vernommen werden und ebenso nimmt die Varianz im Verlauf zu. Bei dem PERCLOS-Merkmal in Abbildung 6.8 wird der Anstieg im Mittel über die Zeit noch etwas deutlicher, aber die Varianz steigt hier nicht so sehr an. Gleiches gilt für das Merkmal Blinzelfrequenz, welches in Abbildung 6.9 abgebildet ist. Um hier auf einen Unterschied zu dem PERCLOS-Merkmal zu verweisen, kann der Zeitraum zwischen Minute 20 und 25 herangezogen werden. Hier ist ein Abfall im Mittel des PERCLOS-Merkmals zu erkennen, wohingegen das Merkmal Blinzelfrequenz in diesem Zeitinterval, abgesehen von der Varianz, im Mittel nicht abfallend wirkt. Um eine Vorstellung von der Form der Sitzdruckdaten zu bekommen, ist beispielhaft das Merkmal SP6 in Abbildung 6.10 dargestellt. Größere Bewegungen des Probanden sind hier zum Beispiel in Minute 12 zu sehen. Nach dieser Bewegung scheint die Sitzposition leicht anders zu sein, sodass der Mittelwert daraufhin etwas tiefer ist. Die Anzahl der versteckten Einheiten ist bei diesem Modell mit 150 halb so groß wie bei der Vorstudie. Der restliche Ablauf im Netzwerk folgt analog zu der Vorgehensweise aus der Vorstudie. Auch hier sind alle Merkmale für eine Klassifikation notwendig, da hier die Fusion vor der Klassifikation stattfindet. In diesem Fall wird das Training über eine bestimmte Anzahl von 600 Epochen durchgeführt. Diese Zahl ist so groß gewählt, dass auch in diesem Fall die Änderung des Fehlers sehr klein ist. Um die Trainingsphase noch besser beurteilen zu können, werden bei der Hauptstudie Validierungsdaten während des Trainings verwendet. Die Auswertung hierzu ist in Abbildung 6.12 dargestellt. In dem oberen Teil der Abbildung ist die Genauigkeit sowohl der Trainingsdaten als auch der Validierungsdaten über die Epochen dargestellt. Die Genauigkeit G in Prozent ([%]) berechnet sich hierbei aus der Anzahl der richtig-positiven Klassifikationen n_{tp}, der Anzahl der richtig-negativen Klassifikationen n_{tn} und der Anzahl aller Klassifikationen n_{ges}:

$$G[\%] = \frac{n_{\text{tp}} + n_{\text{tn}}}{n_{\text{ges}}} \tag{6.2.1}$$

Diese Definition wird auch später für die Auswertung der Testdaten herangezogen. Im unteren Teil ist der Wert der Fehlerfunktion für die Trainings- sowie Validierungsdaten aufgetragen. Die Validierungsdaten werden nur alle 50 Epochen ausgewertet, wohingegen die Trainingsdaten in jeder Epoche evaluiert werden. Bei der Trainingsgenauigkeit wird deutlich, dass schon nach etwa 150 Epochen eine hohe Genauigkeit von über 98 % erreicht werden kann. Diese Epoche ist zudem der letzte Zeitpunkt des Trainings, an dem die Validierungsgenauigkeit steigt und der zugehörige Fehler sinkt. Danach wird für die Validierungsdaten eine ähnliche Genauigkeit nur noch mit einem höheren Fehler erreicht. Dabei steigt die Genauigkeit der Validierungsdaten nicht über 53 %. Dieses Ergebnis deutet darauf hin, dass in diesem Fall eher eine Überanpassung auf die Trainingsdaten vorliegt. Genau aus diesem Grund wurde die Struktur des Netzwerks für die Daten des Hauptversuchs mit der DropOut-Schicht angepasst. Hiermit sollte die Modellgröße reduziert werden, um einer Überanpassung entgegenzuwirken. Diese Maßnahme hat jedoch wenig bis keine Auswirkung auf die im Training beobachtete Problematik. Wie in der Abbildung 6.12 zu sehen ist, verbessert oder verschlechtert sich die Genauigkeit der Valdierungsdaten nicht im größeren Maße, sondern schwankt zwischen 40 % und

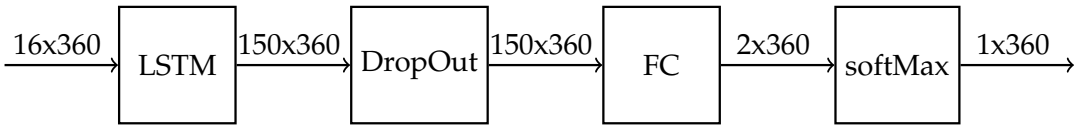


Abbildung 6.11.: Struktur des künstlichen neuronalen Netzwerks mit der Fusion auf Merkmalsebene für die Daten aus dem Hauptversuch

53 %, sodass ein früheres Stoppen des Trainings ebenfalls nicht hilfreich wäre.

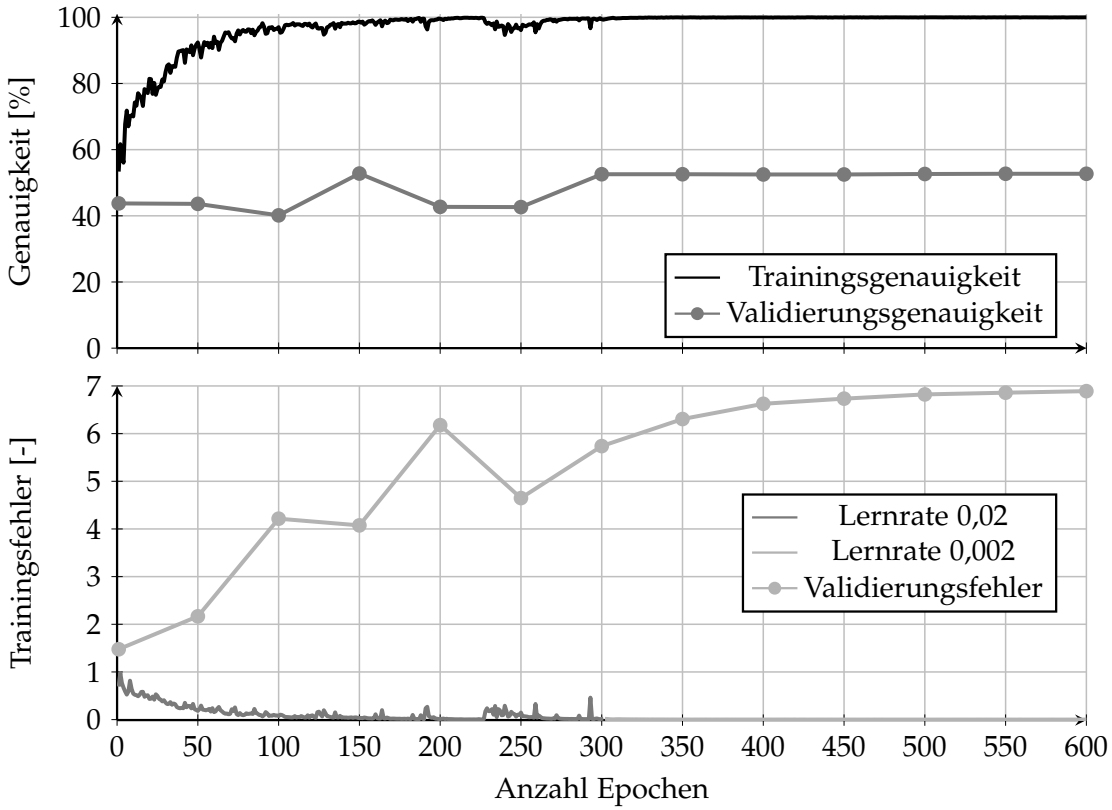


Abbildung 6.12.: Genauigkeit beim Training und der Validierung sowie Wert der Fehlerfunktion beim Training über die Anzahl der Epochen mit dem Datensatz des Hauptversuchs

6.3. Fusion der physiologischen Daten auf Entscheidungsebene

In diesem Abschnitt soll die Struktur der Modelle, die eine Fusion auf Entscheidungsebene durchführen, erläutert werden. Der größte Unterschied zu den Modellen, die auf Merkmalsebene fusionieren, liegt darin, dass hier für jeden Sensor ein Teilmodell mit Klassifikator trainiert wird. Die einzelnen Teilmodelle besitzen dabei die gleiche Struktur wie die im vorherigen Abschnitt vorgestellten Modelle, nur mit einer kleineren Menge an Daten am Eingang. Die Parameter im Training werden ebenfalls beibehalten, um hier eine Aussagekraft für einen Vergleich

Tabelle 6.2.: Merkmale für das Training der Datenverarbeitung des Hauptversuchs

	Körperteil	Merkmal
1	Herz	Herzfrequenz HF
2		Herzraten-Varaiabilität H
3	Augen	PERCLOS
4		Blinzel-Häufigkeit
5	Sitzdruck	SP1
6		SP2
7		SP3
8		SP4
9		SP5
10		SP6
11		SP7
12		SP8
13		SP9
14		SP10
15		SP11
16		SP12

zu schaffen. Dabei wird auch hier wieder sowohl das Modell für die Daten aus dem Vorversuch als auch das Modell für die Daten aus dem Hauptversuch vorgestellt. Da im Hauptversuch mehr Sensoren hinzukommen, ändert sich auch die Anzahl der Teilmodelle der Klassifikation.

Vorversuch

In Abbildung 6.13 ist das Modell für den Ansatz der Fusion auf Entscheidungsebene abgebildet, welches mithilfe der Daten aus der Vorstudie erzeugt wurde. Es besteht aus zwei Teilmodellen, eines für die Daten aus dem Kamerasensor und eines für die Daten aus dem Herzsensor.
Unter dem Begriff Teilmodell ist ein Netzwerk zu verstehen, welches die gleiche Struktur wie im vorherigen Abschnitt in Abbildung 6.4 aufweist, mit dem Unterschied, dass in diesem Fall nur die Merkmale einer der Sensoren für das Training dieses Modells verwendet werden. Bei dem Teilmodell für das Herz werden die Merkmale

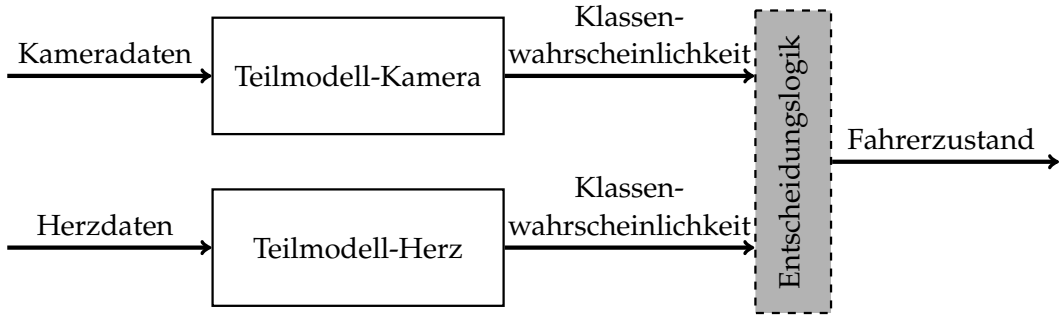


Abbildung 6.13.: Struktur des Modells mit Fusion auf Entscheidungsebene für die Daten des Vorversuchs

HF und *H* verwendet.

Die Entscheidungslogik, welche nach den Teilmodellen folgt, wertet die Klassenwahrscheinlichkeiten aus den Teilmodellen aus. Jedes der Teilmodelle stellt pro Klasse eine Klassenwahrscheinlichkeit zur Verfügung. Da in diesem Fall eine Klassifikation mit zwei Klassen stattfindet, fallen die beiden Klassenwahrscheinlichkeiten symmetrisch aus. Werden die Wahrscheinlichkeiten der unterschiedlichen Klassen von einem Teilmodell zusammenaddiert, muss sich ein Wert von 100 % ergeben. Aus diesem Grund genügt es einen der beiden Werte für eine Auswertung zu verwenden, da in dem anderen Wert keine zusätzliche Information enthalten ist. Die wohl einfachste Möglichkeit, die Daten aus den verschiedenen Teilmodellen zu verwenden, ist das arithmetische Mittel der Klassenwahrscheinlichkeitswerte. So wird beispielsweise der Klassenwahrscheinlichkeitswert der Klasse Hypovigilant von beiden Modellen zusammenaddiert und durch die Anzahl der Teilmodelle geteilt. Bei einem resultierenden gemittelten Klassenwahrscheinlichkeitswert von 50 % wird von einem Zustandswechsel ausgegangen. Hierbei gibt dann das Modell den jeweils wahrscheinlicheren Wert als prädizierte Klasse an.

Das Training wird in dieser Auswertung wie in Abschnitt 6.2 für die Daten aus dem Vorversuch über die gleiche Anzahl von Epochen mit den gleichen Parametern ausgeführt. Jedoch wird beim Training lediglich eine Genauigkeit von etwa 83 % erreicht. Diese liegt somit 15 Prozentpunkte unter dem Wert des Modells, welches mithilfe aller Merkmale trainiert wurde. Dies schlägt sich auch im Wert der Fehlerfunktion nieder, die in Abbildung 6.14 zu sehen ist. Hier ist zum Vergleich der Wert aus Abbildung 6.5 von der Auswertung der Vorstudie integriert worden. Der Wert des vollständigen Modells liegt dabei, ab etwa Epoche 15, unter dem Wert des Teilmodells. Dieses erreicht bis zum Ende keinen Wert, der unter 0,227 liegt. Im Gegensatz dazu sinkt der Trainingsfehler des vollständigen Modells unter 0,003.

Die Ergebnisse des Trainings des zweiten Teilmodells, welches nur mit Daten aus der Kamera trainiert wurde, ist in Abbildung 6.15 zu sehen. Hier wird ebenso deutlich, dass der Trainingsfehler auf ähnlichem Niveau bleibt. In diesem Fall wird kein Wert unter einem Fehler von 0,293 erreicht. Damit liegt der Wert etwa 0,290 höher als bei dem Modell mit allen Merkmalen. Dabei kommt dieses Modell mit den Trainingsdaten auf eine Genauigkeit von etwa 90 %.

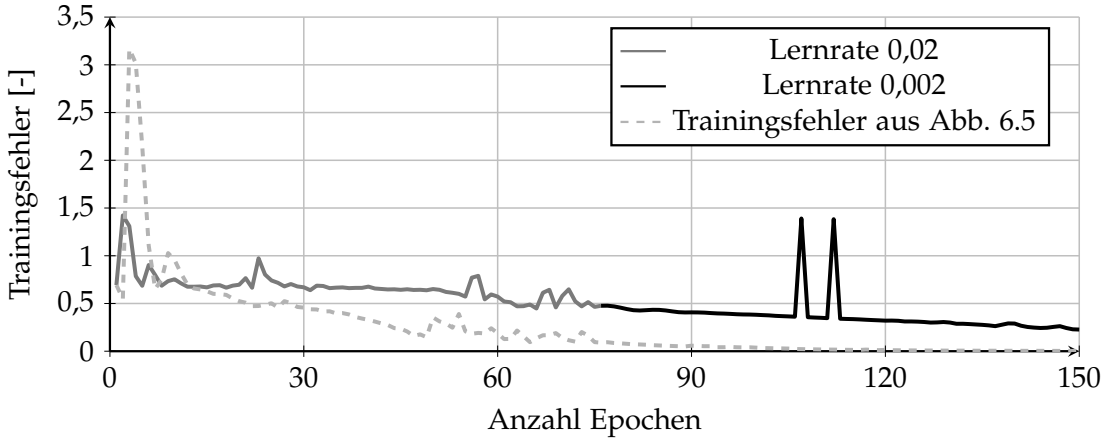


Abbildung 6.14.: Wert der Fehlerfunktion beim Training mit zwei Merkmalen des Herzsensors gegenüber dem Modell mit allen Merkmalen

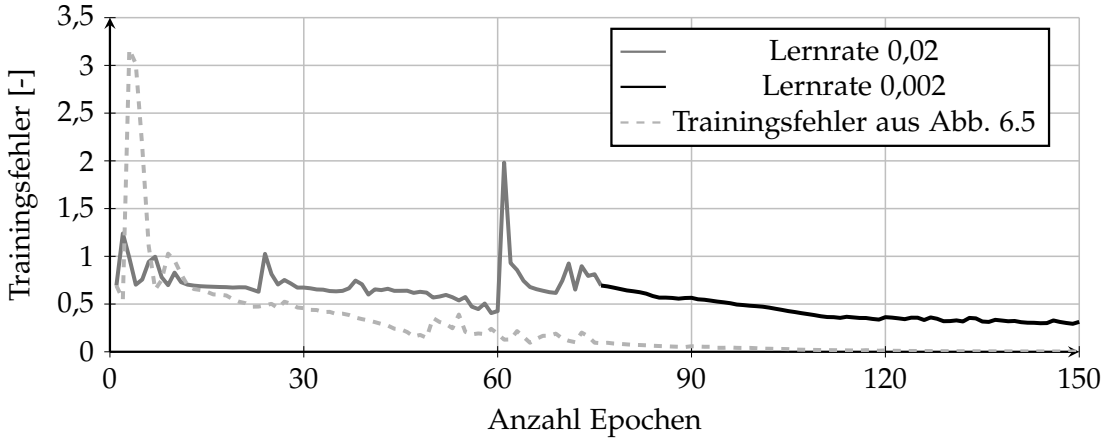


Abbildung 6.15.: Wert der Fehlerfunktion beim Training mit einem Merkmal der Kamera auf den Trainingsdaten von acht Probanden über der Anzahl der trainierten Epochen

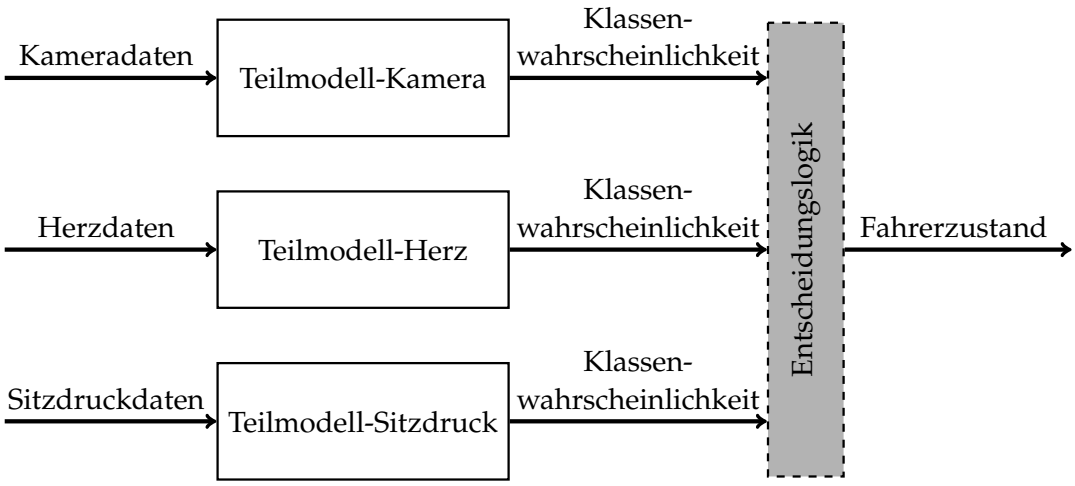


Abbildung 6.16.: Struktur des Modells mit Fusion auf Entscheidungsebene

Hauptversuch

In Abbildung 6.16 ist das Modell des Ansatzes der Fusion auf Entscheidungsebene, für die Daten aus dem Hauptversuch, zu sehen. In diesem Fall kommt ein Teilmodell hinzu, denn bei diesem Versuch wurden die Sitzdrucksensoren ergänzt. Bei dem Teilmodell für die Kameradaten wird wieder ein Merkmal mehr als in dem Modell für die Daten aus dem Vorversuch verwendet, denn im Hauptversuch wird zusätzlich die Blinzel-Häufigkeit bestimmt. Die Teilmodelle haben dabei eine analoge Struktur zu der im vorherigen Abschnitt vorgestellten Form. Für die Daten des Hauptversuchs wird wie zuvor beim Modell aller Merkmalsdaten die Dropout-Schicht ergänzt. In diesem Fall ist wieder eine Entscheidungslogik nach den Teilmodellen angeknüpft. Ebenso wie zuvor beschrieben, wird das arithmetische Mittel aus den Klassenwahrscheinlichkeiten als Entscheidungslogik verwendet, sodass die Ergebnisse vergleichbar bleiben. Bei diesem Modell wird der Wahrscheinlichkeitswert für eine Klasse aus den drei einzelnen Klassenwahrscheinlichkeiten bestimmt.

Das Training der Teilmodelle wird analog zu den vorherigen Modellen aus dem Hauptversuch durchgeführt. So wird jedes Teilmodell über 600 Epochen trainiert. Eine Auswertung der Validierungsdaten findet alle 50 Epochen statt. Die Ergebnisse der Trainings der Teilmodelle sind in Abbildung 6.17 zu sehen. Hierbei werden vier relevante Informationen jedes Teilmodells mit dem Modell aus Abschnitt 6.2 verglichen, welches in der Abbildung mit "Alle" und durch eine gestrichelte Linie in dunkelgrau gekennzeichnet ist. Zu den vier Informationen gehören die Trainingsgenauigkeit, die Validierungsgenauigkeit, der Validierungsfehler und der Trainingsfehler. Das Ergebnis des Trainings vom Modell der Sitzdruckdaten ist in hellgrau, vom Modell der Kameradaten in dunkelgrau und vom Modell der Herzdaten in schwarz dargestellt. Werden die Trainingsgenauigkeiten miteinander verglichen, kann festgestellt werden, dass zum Ende des Trainings alle Modelle, außer das Modell der Herzdaten, einen hohen Wert von mehr als 99 % erreichen. Dahingegen liegt der maximale Wert der Trainingsgenauigkeit des Modells der Sitzdruckdaten sechs Prozentpunkte darunter und bleibt somit unter 93 %. Dafür ist

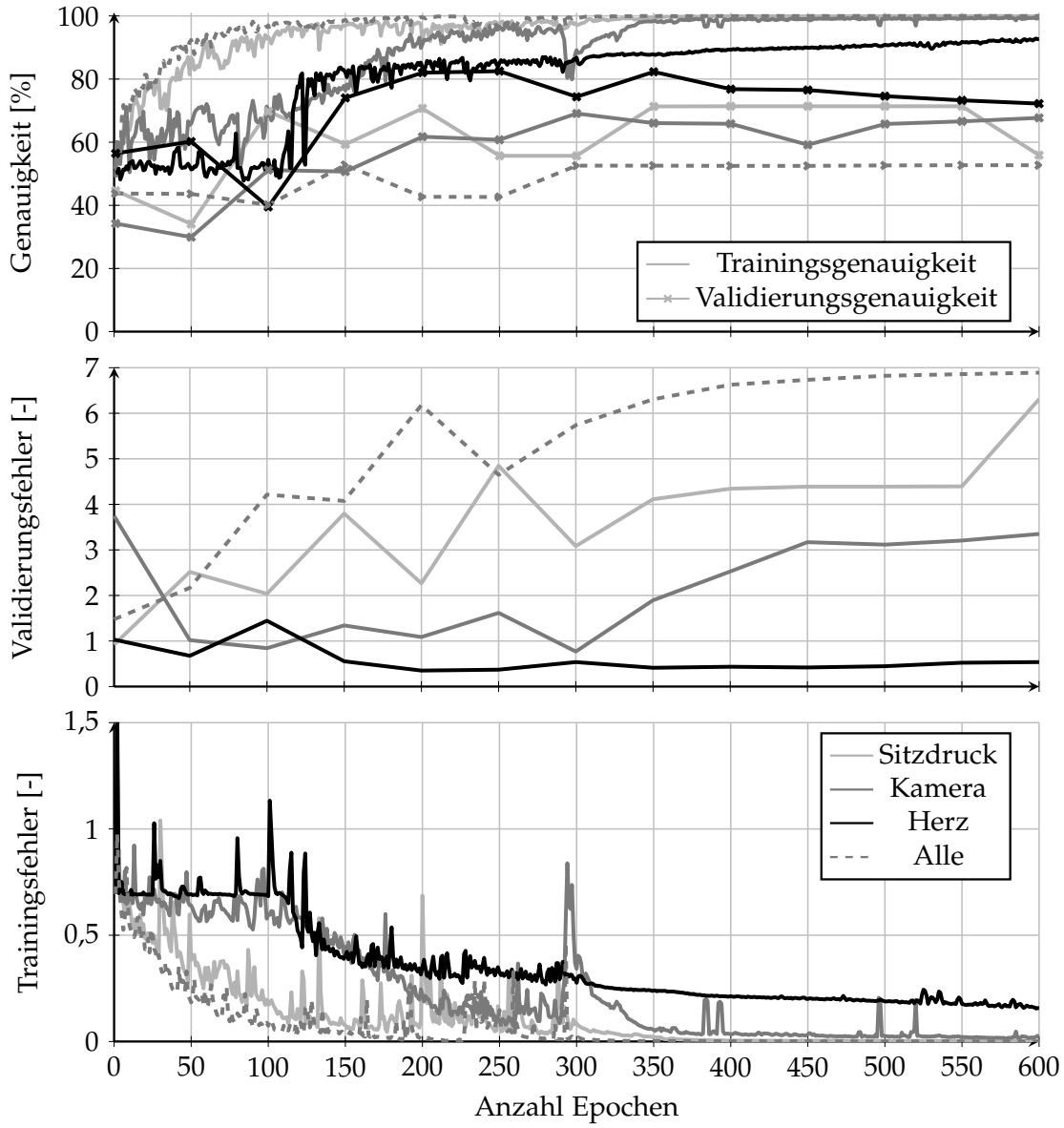


Abbildung 6.17.: Vergleich aller Teilmodelle in Trainings- sowie Validierungsgenauigkeit und Trainings- sowie Validierungs-Fehler

bei diesem Modell die Validierungsgenauigkeit mit etwa 72 % am höchsten unter den Modellen, welche aber ab Epoche 350 von einem Wert von etwa 82 % stetig fällt. Alle anderen Modelle sind in der Validierungsgenauigkeit schlechter, jedoch fast in keinem Fall schlechter als das Modell, welches mit allen Sensordaten trainiert wurde. Dieses Modell erreicht bei der Validierungsgenauigkeit keinen Wert, der über 53 % liegt. Bei dem Validierungsfehler zeichnet sich ein ähnliches Bild ab. Der größte Fehler wird von dem Modell aller Daten erzeugt. Das Teilmodell der Sitzdruckdaten hat einen etwas kleineren Fehler. Darauf folgt das Teilmodell der Kameradaten mit dem zweitgeringsten Fehler. Mit dem geringsten Fehler steht das Modell der Herzdaten dar. Ab etwa Epoche 300 wird deutlich, dass der Validierungsfehler bei allen Modellen, außer beim Sitzdruckdatenmodell, gegen einen bestimmten Wert konvergiert. Die Varianz in dem Validierungsfehlersignal ist hier geringer, da die Lernrate auch bei diesen Trainings ab Epoche 300 von 0,02 auf 0,002 verringert wird. Nur der Validierungsfehler des Modells aus den Sitzdruckdaten steigt in der letzten Auswertung stark an. Dies zeigt sich auch in der Validierungsgenauigkeit, die mit diesem Schritt schlechter als die Valdierunggenauigkeit des Modells der Kameradaten wird. Der Trainingsfehler läuft bei allen Modellen, bis auf das Modell der Herzdaten, gegen einen Wert von nahezu Null. Das Modell der Herzdaten erreicht hierbei nur einen Wert von etwa 0,16, wohingegen der Fehler aller anderer Modelle unter einem Wert von 0,02 liegen. Auch hier nimmt die Varianz in den Signalen ab Epoche 300 wieder ab, was mit der verringerten Lernrate zusammenhängt.

7

Modell zur Klassifikation von Fahreraufmerksamkeit mit Fokus auf Vigilanz

In diesem Kapitel geht es um die Auswertung der zuvor vorgestellten Modelle auf ungesehenen Testdaten und den Vergleich unter den verschiedenen Fusionsmethoden. Dabei wird zudem zwischen dem Vorversuch und dem Hauptversuch unterschieden. Dies ist notwendig, da sich die Datensätze aus den Versuchen in der Anzahl der Merkmale unterscheiden und die Modelle eine geringfügig andere Struktur besitzen. Im zweiten Teil dieses Kapitels wird eine Einschätzung formuliert, welches der genannten Modelle für den Einsatz in einem Fahrzeug am besten geeignet scheint.

7.1. Performanz der verschiedenen Fusionsmethoden

Die Darstellung der Performanz der verschiedenen Fusionsmethoden erfolgt durch die Auswertung von Daten, die beim Training nicht mit eingeflossen sind. Für die Vorstudie wird ein Testdatensatz, bestehend aus den Daten eines Probanden, ausgewertet. Für die Hauptstudie umfasst der Testdatensatz die Daten von drei Probanden. Dabei werden verschiedene Kriterien zur Ermittlung der Performanz herangezogen, die im Folgenden erklärt werden.

Auswertungskriterien

Für die Bewertung der Modelle gibt es verschiedene Kriterien. In der hier vorliegenden Arbeit werden für die Auswertung die Genauigkeit (siehe Gleichung 6.2.1), die Verzugszeit, die Präzision und die Ausfallquote als Kriterien herangezogen. Die Verzugszeit ist ein spezielles Kriterium, welches im Gesamtkontext, der vorliegenden Arbeit, eingeordnet werden sollte. Dabei geht es darum, wie viel Zeit zwischen einer Vorhersage des Modells und der tatsächlich wahren Änderung im Zustand vergeht. Da in den Testdaten nur ein Zustandswechsel vorliegt, kann der Zeitpunkt der Vorhersage des Modells mit t_{m} und der Zeitpunkt der wahren Änderung mit t_{tp}

angegeben werden. Damit ergibt sich die Verzugszeit t_v mit

$$t_v[s] = |t_m - t_{tp}|. \tag{7.1.1}$$

Die Präzision P ergibt sich aus der Anzahl der richtig-positiven Klassifikationen n_{tp} und der Anzahl der falsch-positiven Klassifikationen n_{fp} wie folgt:

$$P[\%] = \frac{n_{tp}}{n_{tp} + n_{fp}}. \tag{7.1.2}$$

Die Ausfallquote R (auch als Sensitivität bekannt) ergibt sich aus der Anzahl der richtig-positiven Klassifikationen n_{tp} und den falsch-negativen Klassifikationen n_{fn} nach folgender Formel:

$$R[\%] = \frac{n_{tp}}{n_{tp} + n_{fn}} \tag{7.1.3}$$

Die Spezifität S wird aus der Anzahl der richtig-negativen Klassifikationen n_{tn} sowie der Anzahl der falsch-positiven Klassifikationen n_{fp} mit folgender Formel berechnet:

$$S[\%] = \frac{n_{tn}}{n_{tn} + n_{fp}} \tag{7.1.4}$$

Dazu kann noch das Maß des negativen Vorhersagewerts V herangezogen werden, welches durch die Anzahl der richtig-negativen Klassifikationen n_{tn} und der Anzahl der falsch-negativen Klassifikationen n_{fn} wie folgt berechnet wird:

$$V[\%] = \frac{n_{tn}}{n_{tn} + n_{fn}} \tag{7.1.5}$$

Da in der hier vorliegenden Arbeit eine binäre Klassifikation vorliegt und nur ein Zustandswechsel in den Daten prädiziert werden soll, können einige der oben genannten Bewertungskriterien ineinander überführt werden oder treffen lediglich triviale Aussagen. So lässt sich die Verzugszeit nur mit einer Auflösung von zehn Sekunden bestimmen, da das Netzwerk ebenfalls diese zeitliche Auflösung besitzt. Eine Verzugszeit t_v von null Sekunden entspricht einer Genauigkeit G von 100 %. Eine Verzugszeit t_v von zehn Sekunden im Hauptversuch entspricht einer Genauigkeit von 99,72 %. Jede weitere Abweichung in der Verzugszeit t_v im nächsten diskreten Schritt würde auch einen diskreten Schritt in der Genauigkeit G von 0,28 Prozentpunkten (also $\frac{1}{360}$) im Hauptversuch bedeuten. Da der Zustandswechsel bei einer nicht idealen Prädiktion entweder vor oder nach dem wirklichen Zustandswechsel klassifiziert werden kann, ist entweder $n_{np} = 0$ oder $n_{fp} = 0$, sodass entweder P oder R auf 100 % ansteigen.
Ein Problem, welches bei der Genauigkeit G entsteht, ist die Abhängigkeit des Testsignal. Da in der vorliegenden Arbeit die Zustandswechsel nicht mittig in der Versuchszeit liegen, ist es für das Netzwerk leicht möglich, eine Genauigkeit über 50 % zu erreichen, obwohl es keinen Zustandswechsel prädiziert. So ist die Genauigkeit G für die in der Vorstudie verwendeten Testdaten bereits bei 80 %, obwohl das Netzwerk keinen Zustandswechsel prädiziert.
Dahingegen ist ein Maß, wie die Verzugszeit, besser geeignet. Diese wird relativ zu der Einordnung der Daten berechnet und ist somit nicht so sehr vom Testsignal abhängig. Viele dieser Bewertungskriterien lassen sich in einer Konfusionsmatrix zusammenfassen, wie es in Tabelle 7.1 zu sehen ist.

Tabelle 7.1.: Konfusionsmatrix Erklärung

		Berechnet		
		Hypovigilant	Vigilant	
Vorgabe	Hypovigilant	n_{tp}	n_{fn}	Ausfallquote R
	Vigilant	n_{fp}	n_{tn}	Spezifität S
		Präzision P	negativer Vorhersagewert V	Genauigkeit G

Ergebnisse der Vorstudie

Nachfolgend werden nun die Ergebnisse der Netzwerke mit der Struktur aus Abbildung 6.4 gezeigt und erläutert.
In Abbildung 7.1 ist das Ergebnis des Netzwerks zu sehen, welches alle drei Merkmale (aus Herz und Auge) für das Training verwendet und auf der Fusion der Merkmalsebene basiert. Es ist nicht sinnvoll, an dieser Stelle Merkmale aus der Gehirnaktivität zu benutzen, da diese bereits bei der Einordnung der Daten verwendet wurden. Bei den Testdaten kann eine Genauigkeit von 96 % festgestellt werden. Damit ist das Netzwerk um 31 Prozentpunkte besser als ein einfaches vollständig verknüpftes Netzwerk. Der Zeitverzug zwischen den Zustandswechseln beträgt 100 s. Die Präzision liegt bei 100 % und die Ausfallquote bei 81,48 %. Dazu gehen aus der Tabelle 7.2 eine Spezifität von 100 % und ein negativer Vorhersagewert von etwa 95 % hervor. Somit wird der Unterschied zum einfachen Netzwerk deutlich, denn bei diesem Netzwerk mit Gedächtnis variiert der Zustand nicht, sodass die Präzision deutlich besser ist. Um die Robustheit des Verfahrens zu testen, wird eine Kreuz-Validierung durchgeführt. Dazu werden Test- und Trainingsdaten mehrfach verändert, indem immer ein anderer Proband als Testdatensatz fungiert und die übrigen acht als Trainingsdatensatz. Dies wird für alle neun Probanden wiederholt, sodass jeder Proband einmal als Testdatensatz herangezogen wird. Hierbei erreicht dieses Netzwerk eine Genauigkeit von ∅86,2 %, die sich nur um etwa 10 Prozentpunkte von der Auswertung des einzelnen Testdatensatzes verringert.

Ergebnisse der Hauptstudie

Da in der Hauptstudie eine größere Anzahl von Probanden untersucht wurde, können hier mehrere Probanden als Testdaten herangezogen werden. In diesem Fall werden die Daten von drei Probanden ausgewertet. Für diese Probanden wird folglich die Abkürzung T_i mit $i \in [1,3]$ verwendet. Für das Modell aus der Abbildung 6.11 sind die Ergebnisse in der Abbildung 7.2 dargestellt. Hier wird für jeden der drei Testdatensätze die berechnete Ausgabe des Netzwerks und der durch die Fragebögen vorgegebene Zustandswechsel dargestellt. In der Abbildung wird deutlich, dass sich die vorgegebenen Zustandswechsel sehr unterscheiden können.

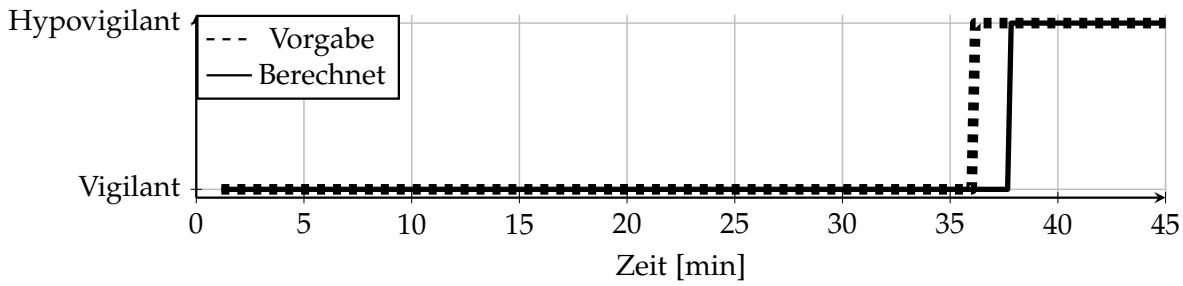


Abbildung 7.1.: Klassifikation auf einer Sequenz der Testdaten mit allen verfügbaren Merkmalen auf Basis der Fusion auf Merkmalsebene mit den Testdaten des Hauptversuchs

Tabelle 7.2.: Konfusionsmatrix Modell mit Fusion auf Merkmalsebene mit den Testdaten aus dem Vorversuch

		Berechnet		
		Hypovigilant	Vigilant	
Vorgabe	Hypovigilant	44	10	$R = 81,48\,\%$
	Vigilant	0	210	$S = 100,00\,\%$
		$P = 100,00\,\%$	$V = 95,45\,\%$	$G = 96,21\,\%$

So tritt dieser bei Datensatz T_1 gegenüber den anderen beiden Datensätzen erst in Minute 42 auf. Bei diesem Datensatz wird zudem deutlich, dass das Netzwerk gar keinen Zustandswechsel berechnet hat, wobei eine Genauigkeit G von etwa 69 % resultiert. Bei Datensatz T_2 und T_3 berechnet das Netzwerk hingegen einen Zustandswechsel für die Daten. Dies funktioniert für T_2 am besten mit einer Genauigkeit von etwa 99 %. Für den Datensatz T_3 ergibt sich eine Genauigkeit, die um etwa 16 Prozentpunkte schlechter ist als bei dem Datensatz T_2. Diese ist mit etwa 84 % jedoch um 15 Prozentpunkte besser als bei dem Datensatz T_1. Wenn nun die Verzugszeit t_v ausgewertet wird, zeigt sich ein ähnliches Bild. Mit dem Datensatz T_1 ist keine Verzugszeit anzugeben, da hier das Netzwerk keinen Zustandswechsel berechnet. Bei dem Datensatz von T_2 ist eine sehr geringe Verzugszeit t_v von nur etwa 30 Sekunden erreicht worden. Der Datensatz T_3 bringt eine um 560 Sekunden größere Verzugszeit t_v hervor, welche somit 590 Sekunden beträgt. Die Auswertung von Präzision P und Ausfallquote R ergibt in diesem Fall, dass beim Datensatz T_1 keine Angabe zur Präzision P gemacht werden kann, da hier weder richtig-positive noch falsch-positive Klassifikationen auftreten. Der Wert für die Ausfallquote R liegt mit diesen Daten bei 0 %. Dies ist wieder auf die null richtig-positiven Klassifikationen zurückzuführen. Beim Datensatz T_2 hingegen kann der Präzisionswert P mit 100 % und die Ausfallquote mit 99 % angegeben werden. Diese Werte sind auf die nur wenigen Fehlklassifikationen in drei Zeitschritten zurückzuführen, welche ausschließlich falsch-negative Klassifikationen sind. Dies ist beim Datensatz T_3 genau

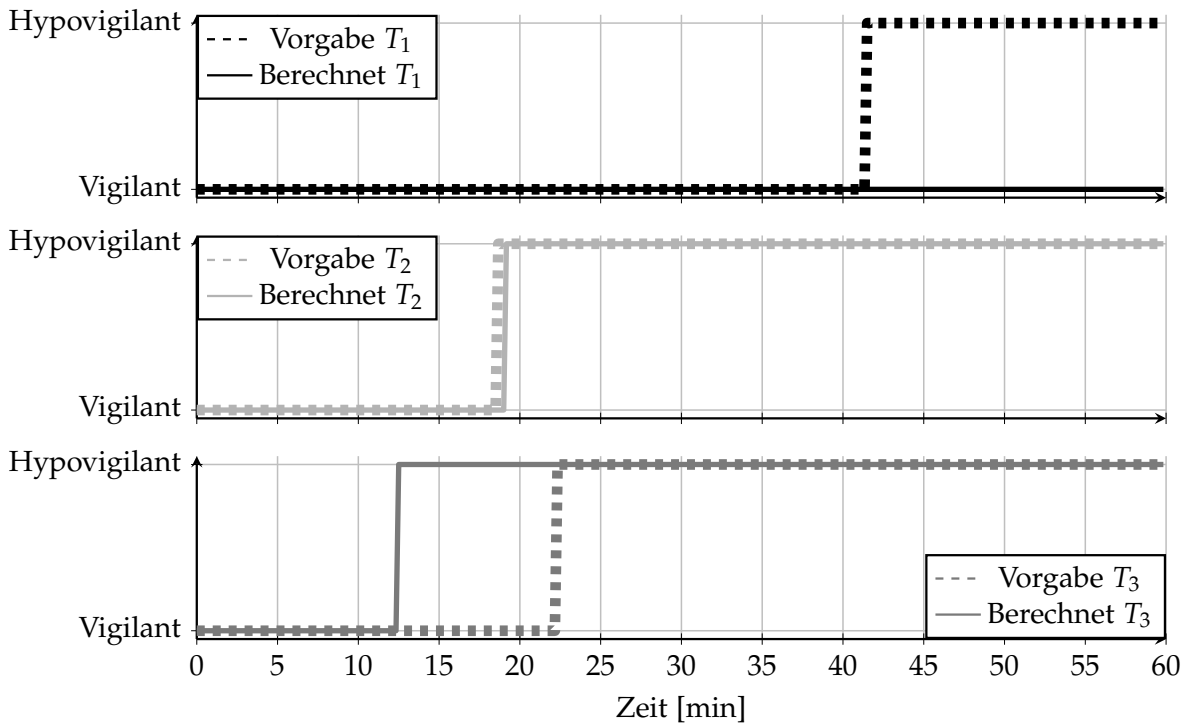


Abbildung 7.2.: Klassifikation der Testdaten von drei Probanden des Hauptversuchs mit allen verfügbaren Merkmalen mit dem Modell auf Basis der Fusion auf Merkmalsebene

umgekehrt, hier gibt es nur falsch-positive Klassifikationen. So verändern sich auch die Werte für die Ausfallquote R zu 100 %, welche somit um einen Prozentpunkt höher ist, und für die Präzision P zu 79 %, also 31 Prozentpunkte weniger als beim Datensatz T_2.
Aus Tabelle 7.3 geht für den Testdatensatz $T1$ zusätzlich der Wert der Spezifität von 100 % sowie für den negativen Vorhersagewert von etwa 69 % hervor. Verglichen mit den Werten von dem Testdatensatz $T2$ aus Tabelle 7.4 steht der Testdatensatz beim negativen Vorhersagewert mit etwa 97 % um 28 Prozentpunkte besser da. Die Spezifität ist bei beiden Testdatensätzen identisch. Nur der negative Vorhersagewert des Testdatensatzes $T3$ aus Tabelle 7.5 erreicht mit 100 % einen noch höheren Wert,

Tabelle 7.3.: Konfusionsmatrix Modell mit Fusion auf Merkmalsebene mit dem Testdatensatz $T1$

		Berechnet		
		Hypovigilant	Vigilant	
Vorgabe	Hypovigilant	0	111	$R = 0,00\,\%$
	Vigilant	0	249	$S = 100,00\,\%$
		$P =$ %	$V = 69,17\,\%$	$G = 69,17\,\%$

Tabelle 7.4.: Konfusionsmatrix Modell mit Fusion auf Merkmalsebene mit dem Testdatensatz $T2$

		Berechnet		
		Hypovigilant	Vigilant	
Vorgabe	Hypovigilant	245	3	$R = 98,79\,\%$
	Vigilant	0	112	$S = 100,00\,\%$
		$P = 100,00\,\%$	$V = 97,39\,\%$	$G = 99,17\,\%$

Tabelle 7.5.: Konfusionsmatrix Modell mit Fusion auf Merkmalsebene mit dem Testdatensatz $T3$

		Berechnet		
		Hypovigilant	Vigilant	
Vorgabe	Hypovigilant	226	0	$R = 100,00\,\%$
	Vigilant	59	75	$S = 55,97\,\%$
		$P = 79,30\,\%$	$V = 100,00\,\%$	$G = 83,61\,\%$

jedoch auch mit einem um 44 Prozentpunkte geringeren Wert für die Spezifität mit einem Wert von etwa 56 %.

Ergebnisse der Fusion auf Entscheidungsebene

In diesem Abschnitt der Arbeit sollen nun die Ergebnisse der Fusion auf Entscheidungsebene dargestellt werden. Für diese wird zunächst die Performanz der einzelnen Teilmodelle vorgestellt und danach die Gesamtperformance dieses Ansatzes bestimmt. Dabei werden sowohl die Daten der Vorstudie als auch aus dem Hauptversuch beleuchtet.

Ergebnisse der Vorstudie

Die Ergebnisse des Teilmodells, welches nur mit dem Merkmal der Kameraauswertung (nur ein Merkmal aus den Augen) trainiert wurde, ist in Abbildung 7.3 zu sehen. Hier kann erkannt werden, dass sich das Ergebnis gegenüber dem Modell der Merkmalsfusion verschlechtert und sich mit etwa 83 %, also 13 Prozentpunkte verringert, darstellt. Es wird deutlich, dass dieses Netzwerk den Wechsel des Zustands deutlich zu früh vorhersagt, sodass der Fehler zwischen der Minute 28 und 36 entsteht. Somit lässt sich hier die Verzugszeit t_v zwischen den Zustandswechseln mit 450 s angeben. Diese ist im Vergleich zu dem Netzwerk, das mit mehr Merkmalen aus der Vorstudie trainiert wurde, um 350 s größer. Die Präzision P ist bei einem Wert von

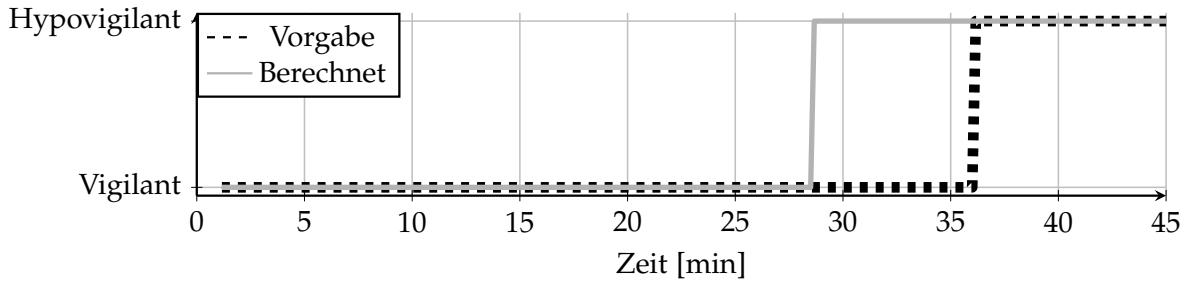


Abbildung 7.3.: Klassifikation auf einer Sequenz der Testdaten des Vorversuchs nur mit dem Teilmodell der Kameradaten

54,55 % und somit um 45 Prozentpunkte schlechter als bei dem Netzwerk, welches mit Merkmalen aus verschiedenen Sensoren gespeist wird. Diese Veränderung ergibt sich durch die zu frühe Klassifikation des Zustandswechsels. Die Ausfallquote R liegt hingegen bei 100 % und ist damit um etwa 24 Prozentpunkte besser als bei dem zuvor gezeigtem Netzwerk. Diese Werte sind stark von dem Testsignal abhängig. Da hier der Zustandswechsel recht spät stattfindet, kann das Netzwerk mit der angegebenen Verzugszeit gar nicht besser in dieser Bewertung werden, obwohl erwartet werden kann, dass der Fahrer noch länger in dem veränderten Zustand bleiben würde. Dennoch wird deutlich, dass der berechnete Zustand nicht springt. Dies ist wichtig, da es kritischer wäre, eine Fehlklassifikation am Ende der Messung zu erhalten.
Ebenso kann die Auswertung für das Teilmodell durchgeführt werden, welches nur mit den Merkmalen vom Herzsensor (also dem Wert der Herzfrequenz HF und dem Wert der Herzraten-Variabilität H) trainiert wird. Das Ergebnis dazu ist in Abbildung 7.4 zu sehen. In diesem Fall wird der Wechsel des Fahrerzustands erst in Minute 40 von dem Netzwerk berechnet, sodass hier eine Genauigkeit von etwa 90 % resultiert. Dieses Ergebnis ist also sechs Prozentpunkte schlechter als das Ergebnis des Modells auf Basis der Fusion auf Merkmalsebene. Jedoch ist es sieben Prozentpunkte besser als das Ergebnis, welches von dem Modell auf Basis der Kameradaten erreicht wurde. Hier ist der zeitliche Versatz 270 s und somit 170 s größer als bei dem Modell, das auf Merkmalsebene fusioniert. Allerdings ist diese Verzugszeit um 180 s geringer als bei dem Netzwerk, welches nur mit den Merkmalen des Kamerasensors trainiert wurde. Die Präzision liegt hierbei bei einem Wert von 100 %, also genauso hoch wie bei dem Netzwerk, welches mit allen Merkmalen trainiert wurde und etwa 55 Prozentpunkte besser als bei dem Netzwerk, welches nur mit Merkmalen des Herzens arbeitet. Dies lässt sich durch die ebenfalls zu späte Klassifikation des Zustandswechsels erklären. So verändert sich auch der Wert der Ausfallquote auf 51,85 %. Dieser Wert ist der schlechteste unter den Netzwerken mit Gedächtnis. Auch hier trägt das Testsignal maßgeblich zu diesem Ergebnis bei. So wird deutlich, dass das berechnete Signal nach dem Zustandswechsel konstant bleibt. Auch ist es wichtig, dass zum Ende das Signal richtig klassifiziert wird, da Fehlklassifikationen zu späten Zeitpunkten schlechter wären.
In Abbildung 7.5 sind die Klassenwahrscheinlichkeiten für die beiden Klassen dargestellt, die von dem Modell, welches ausschließlich Daten vom Kamerasensor verwendet, ausgegeben werden. Wie schon in Kapitel 6.3 beschrieben, ist zu

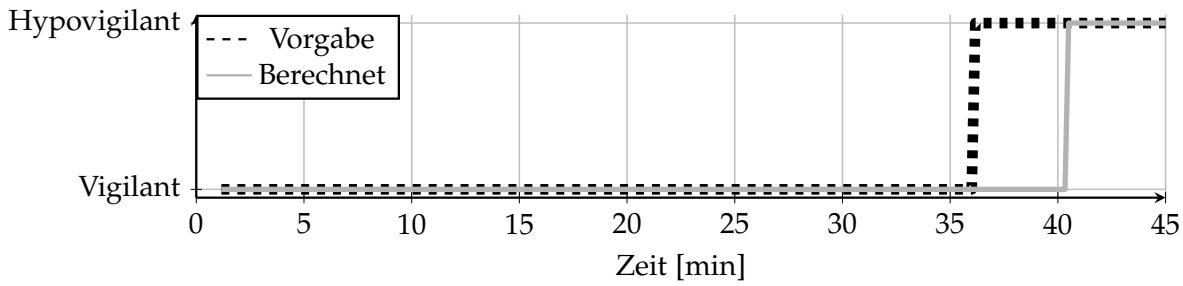


Abbildung 7.4.: Klassifikation auf einer Sequenz der Testdaten des Vorversuchs mit dem Teilmodell der Herzdaten

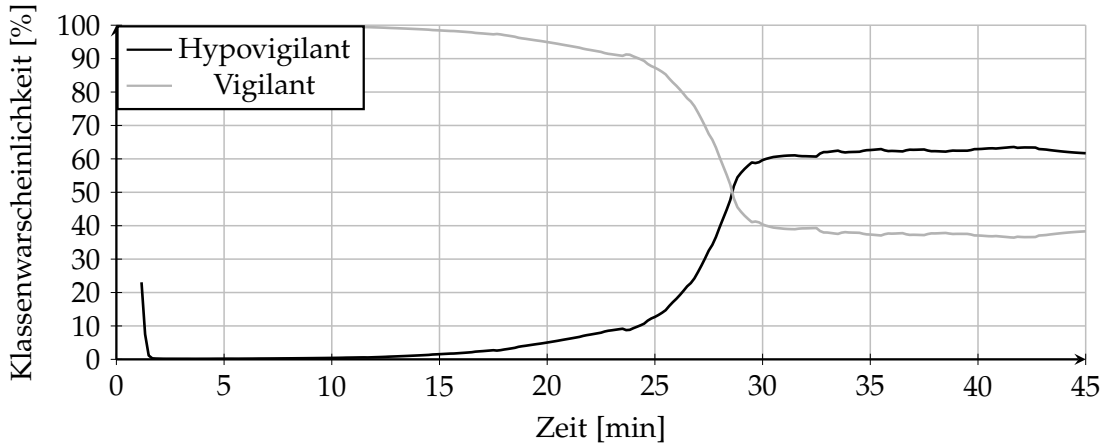


Abbildung 7.5.: Klassenwahrscheinlichkeiten des Teilmodells der Kameradaten auf dem Testdatensatz der Vorstudie

sehen, dass die Klassenwahrscheinlichkeiten der beiden Klassen symmetrisch zueinander sind. Jedoch wird deutlich, dass sich die Klassenwahrscheinlichkeiten über die Zeit hin verändern. In den ersten Minuten ist das Modell noch unsicher, sodass die Klassenwahrscheinlichkeit für die Klasse Hypovigilant in dieser Zeit einer Schwankung von über 20 Prozentpunkten unterliegt. Nach dem Abfall der Klassenwahrscheinlichkeit auf nahezu 0 %, verändert sich diese nur sehr langsam über die nächsten 20 Minuten. Zwischen Minute 20 und 30 ist dann wieder ein größerer Anstieg in der Klassenwahrscheinlichkeit zu erkennen. Nach Minute 30 ist die Klassenwahrscheinlichkeit wieder recht konstant, allerdings bei einem Wert von etwa 60 %. Dies spricht für eine sehr große Unsicherheit des Modells, da der Unterschied zwischen den beiden Klassen bei nur etwa 20 Prozentpunkten liegt.

In Abbildung 7.6 sind die Klassenwahrscheinlichkeiten auf dem Testdatensatz des Vorversuchs gezeigt, die durch das Teilmodell berechnet sind, welches nur mit Daten des Herzsensors trainiert wurde. Hier kann ebenso, wie in der zuvor gezeigten Grafik, in den ersten Minuten eine größere Änderung der Klassenwahrscheinlichkeiten gesehen werden. Von der dritten Minute bis etwa Minute 30 wird keine große Änderung in den Klassenwahrscheinlichkeiten deutlich. Bei etwa Minute 41 ist dann ein deutlicher Anstieg der Klassenwahrscheinlichkeit der Klasse Hypovigilant erkennbar. Analog dazu fällt symmetrisch die Klassenwahrscheinlichkeit für die

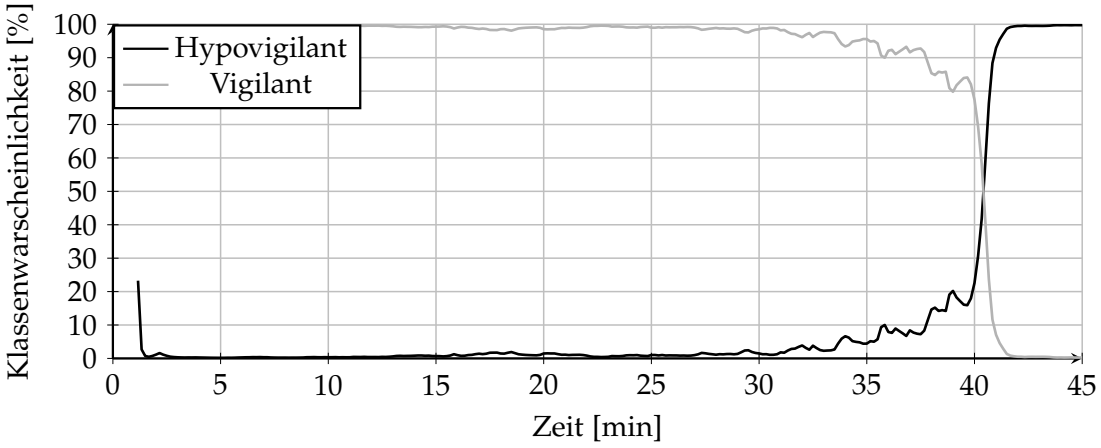


Abbildung 7.6.: Klassenwahrscheinlichkeiten des Teilmodells der Herzdaten auf dem Testdatensatz der Vorstudie

Klasse Vigilant. Nach dem Wechsel der Klassenwahrscheinlichkeit bleiben beide Werte weit auseinander, sodass hier das Netzwerk sehr sicher bei der Einschätzung der Klassen scheint. Es liegen nahezu 100 Prozentpunkte zwischen den beiden Klassenwahrscheinlichkeiten.

In Abbildung 7.7 ist die resultierende Klassenwahrscheinlichkeit, nach der zuvor vorgestellten Fusion, durch Mittlung der beiden Klassenwahrscheinlichkeiten der einzelnen Teilmodelle abgebildet. Auch hier ist wieder in den ersten Minuten eine große Änderung zu sehen. In den darauf folgenden 25 Minuten bleibt das Signal recht konstant unter 10 %. Anschließend wird der Einfluss des Teilmodells aus den Kameradaten deutlich, welcher einen Anstieg um etwa 25 Prozentpunkte bewirkt. Dieser ist jedoch noch nicht für einen Zustandswechsel ausreichend, da die kombinierte Klassenwahrscheinlichkeit hier mit etwa 30 bis 40 % noch zu niedrig ist. Erst knapp nach Minute 40 erreicht die kombinierte Klassenwahrscheinlichkeit einen Wert von über 50 %, sodass dieses Gesamtmodell den Wechsel des Fahrerzustands klassifiziert. Damit erreicht dieses Modell eine Genauigkeit G von etwa 91 %, liegt dabei aber nur 0,38 Prozentpunkte über der Klassifikation des Teilmodells für die Daten vom Herzen. Da sich die Auswertung der Teilmodelle der Herzdaten und Kameradaten im Wesentlichen nicht groß unterscheiden, wird folgend nur noch der Unterschied zu dem Teilmodell der Herzdaten beschrieben. Die Verzugszeit t_v sinkt bei dem fusionierten Modell auf 250 s, was einer Verbesserung von 20 s gegenüber dem Teilmodell der Herzdaten, also zwei Zeitschritten, entspricht. Die Präzision P liegt bei dem Modell nach der Fusion bei 100 %, also genau gleich zu dem Teilmodell der Herzdaten. Dieses Teilmodell schneidet allerdings bei der Ausfallquote um etwa zwei Prozentpunkte schlechter ab, denn das Modell mit Fusion auf Entscheidungsebene erreicht einen Wert von 53,70 %. Dazu geht aus Tabelle 7.6 für das kombinierte Modell mit Fusion auf Entscheidungsebene eine Spezifität von 100 %, sowie ein negativer Vorhersagewert von etwa 89 % hervor.

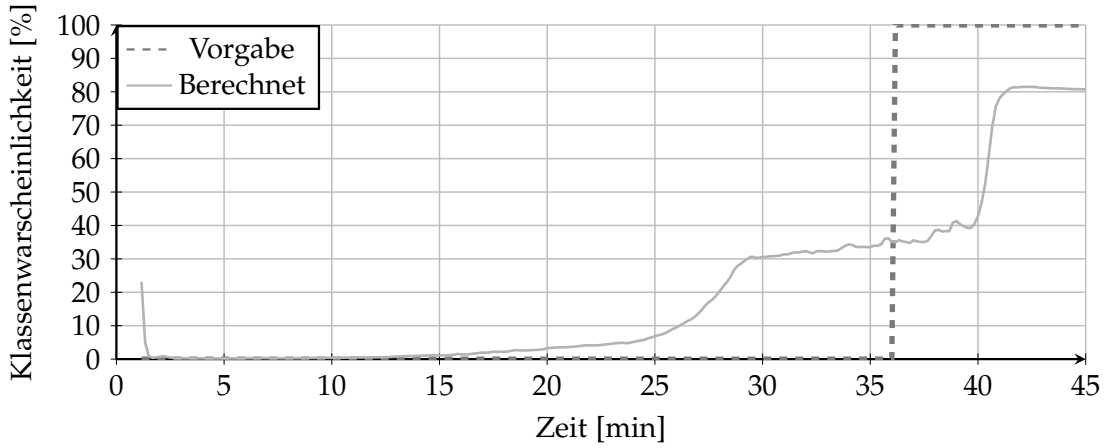


Abbildung 7.7.: Kombination der Klassenwahrscheinlichkeiten des Testdatensatzes der Vorstudie

Tabelle 7.6.: Konfusionsmatrix Modell mit Fusion auf Entscheidungsebene mit den Testdaten aus dem Vorversuch

		Berechnet		
		Hypovigilant	Vigilant	
Vorgabe	Hypovigilant	29	25	$R = 53,70\,\%$
	Vigilant	0	210	$S = 100,00\,\%$
		$P = 100,00\,\%$	$V = 89,36\,\%$	$G = 90,53\,\%$

Ergebnisse der Hauptstudie

Auch in diesem Fall wird zunächst die Performanz der einzelnen Teilmodelle vorgestellt, sodass dann später die Kombination der Teilmodelle besser eingeordnet werden kann. Hierbei umfasst das Teilmodell der Kameradaten nicht nur das Merkmal PERCLOS, sondern auch noch die Blinzel-Häufigkeit. Die Ergebnisse zu den einzelnen Testdatensätzen sind in Abbildung 7.8 zu sehen. Dieses Teilmodell erreicht eine Genauigkeit G von durchschnittlich 90 % und liegt damit um etwa sechs Prozentpunkte über der durchschnittlichen Genauigkeit von etwa 84 % des Modells, welches auf Merkmalsebene fusioniert. Dieser Wert lässt sich dadurch erklären, dass auch bei T_1 ein Zustandswechsel vom Netzwerk berechnet wird. So verbessert sich die Genauigkeit G auf einen Wert von etwa 88 % und ist damit um 19 Prozentpunkte besser als das Modell, welches mit allen Merkmalen trainiert wurde. Wird die Genauigkeit von T_2 der beiden Modelle verglichen, wird klar, dass hier eine Differenz von etwa fünf Prozentpunkten auszumachen ist. Dabei schneidet das gezeigte Modell der Kameradaten mit einer Genauigkeit G von etwa 94 % schlechter ab. Gegenüber dem Modell der Fusion auf Merkmalsebene, lässt sich bei dem Datensatz T_3 mit dem Teilmodell der Kameradaten wieder eine Verbesserung von etwa drei Prozentpunkten auf einen Wert von etwa 87 % beobachten. Wenn es um die Verzugszeit t_v geht, kann bei dem Datensatz T_1 mit dem Teilmodell der Kameradaten eine Zeit von 440 s angegeben werden. Dieser Wert lässt sich nicht vergleichen, da bei dem zuvor vorgestellten Modell auf den Daten des Hauptversuchs keine Zustandsänderung berechnet wurde. Bei dem Datensatz T_2 liegt die Verzugszeit bei dem vorgestellten Teilmodell bei 230 s, welche im Gegensatz zu dem Modell auf Basis von Merkmalsfusion um 200 s abweicht. Hier zeigt sich, dass das Modell, welches über alle Merkmale verfügt, bei einigen Datensätzen besser abschneiden kann. Beim Datensatz T_3 kann das Teilmodell der Kameradaten mit einer Verzugszeit von 460 s das größere Modell um 100 s übertreffen. Ähnlich wie bei der Verzugszeit kann für den Datensatz T_1, bei dem zuvor vorgestellten Modell, kein Wert für die Präzision P angegeben werden, wohingegen das vorgestellte Teilmodell der Kameradaten einen Wert von etwa 72 % hervorbringt. Für den Datensatz T_2 schneiden beide Modelle mit 100 % gleich ab. Bei dem Datensatz T_3 erreicht das Teilmodell wieder einen etwas besseren Wert, denn hier liegt die Präzision P bei einem Wert von etwa 83 % etwa 4 Prozentpunkte über dem Präzisionswert von dem Modell, welches auf Merkmalsebene alle Sensordaten fusioniert. Die Ausfallquote R ist beim Datensatz T_1 konträr zu dem vorher gezeigten Modell, denn bei dem gezeigten Modell der Kameradaten gibt es keine falsch-negativen Klassifikationen, sodass sich ein Wert von 100 % ergibt. Dahingegen übertrifft das Modell, welches mit allen Merkmale trainiert wurde, dieses Modell um neun Prozentpunkte in der Ausfallquote, wenn der Datensatz T_2 verwendet wird. Hierbei erreicht es nur einen Wert von 90 %. Für den Datensatz T_3 zeigt sich bei beiden Modellen ein ähnliches Bild. Da hier keine falsch-negativen Klassifikationen auftreten, ist der Wert für die Ausfallquote in beiden Fällen bei 100 %. Der Vergleich zwischen den Bewertungskriterien (und allen Teilmodellen) kann in Tabelle 7.7 genauer nachvollzogen werden.

In Abbildung 7.9 ist die Performanz des Modells, welches auf den Daten des

Tabelle 7.7.: Zusammenfassung der Ergebnisse der drei Teilmodelle ausgewertet mit den drei Testdatensätze aus dem Hauptversuch

Modell	Kamera				Herz				Sitzdruck			
	G [%]	t_v [s]	P [%]	R [%]	G [%]	t_v [s]	P [%]	R [%]	G [%]	t_v [s]	P [%]	R [%]
T_1	88	440	72	100	55	1620	41	100	75	890	56	100
T_2	94	230	100	91	77	840	100	66	98	70	100	97
T_3	87	460	83	100	84	580	100	74	37	2260	-	0

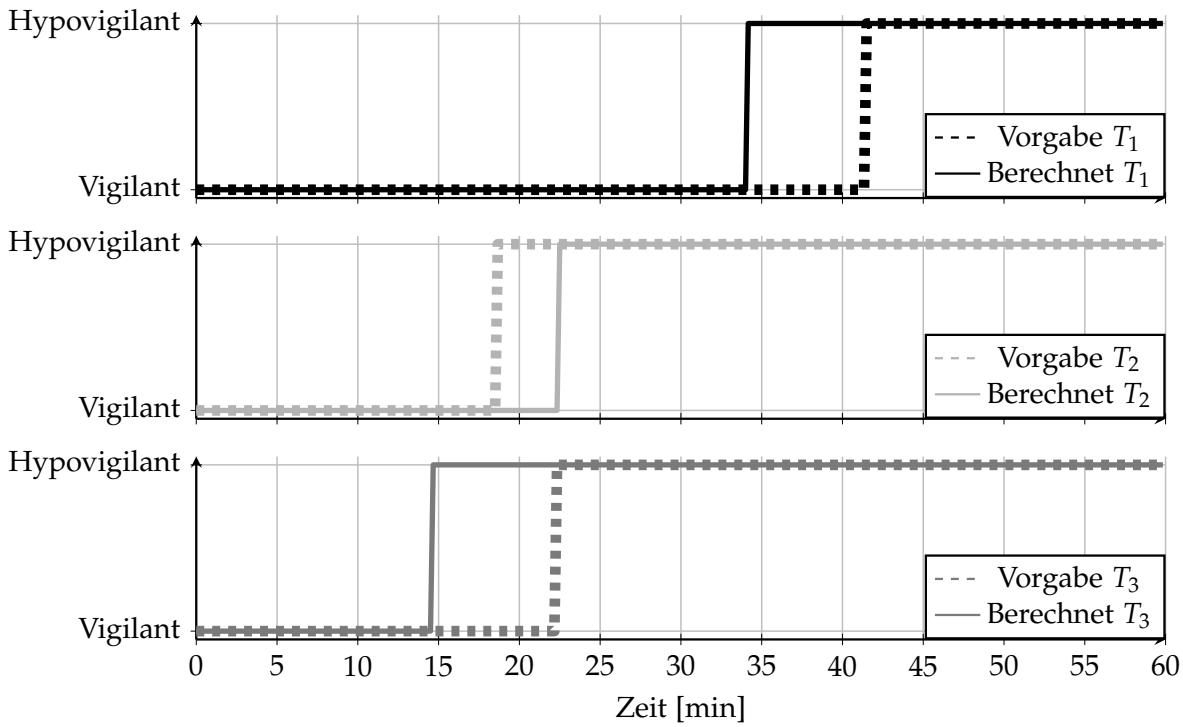


Abbildung 7.8.: Klassifikation der Testdaten von drei Probanden des Hauptversuchs mit dem Teilmodell der Kameradaten

Herzsensors basiert, dargestellt. Auch bei diesem Teilmodell wird deutlich, dass bei dem Testdatensatz T_1 ein Zustandswechsel von dem Modell berechnet wird. In diesem Fall ist der Fehler, der daraus resultiert, jedoch noch höher als bei dem Modell aller Merkmale, welches keinen Zustandswechsel berechnet. Das hier gezeigte Teilmodell erzielt bei diesem Datensatz nur eine Genauigkeit G von 55 % und ist somit um 14 Prozentpunkte schlechter als das Modell, welches keinen Zustandswechsel berechnet. Bei dem Datensatz T_2 zeigt sich ein ähnliches Ergebnis, auch hier besitzt das gezeigte Modell auf Basis der Herzdaten das bisher schlechteste Ergebnis mit einer Genauigkeit G von etwa 77 %. Das Ergebnis ist somit um 17 Prozentpunkte schlechter als das Teilmodell der Kameradaten und 22 Prozentpunkte niedriger als das Modell, welches mit allen Merkmalen aus dem Hauptversuch trainiert wurde. Der Datensatz T_3 hat für das hier gezeigte Modell auf Basis der Herzdaten mit 84 % den gleichen Wert für die Genauigkeit wie das Modell, welches mit allen Merkmalen trainiert wurde. In diesem Fall ist jedoch bei dem Modell der Herzdaten die Prädiktion des Zustandswechsels zu spät und nicht zu früh, wie vorherzusehen war. Diese Information wird sich später in der Beschreibung von Präzision und Ausfallquote niederschlagen. Zunächst werden allerdings die Ergebnisse zur Verzugszeit t_v beschrieben. Diese erreicht bei dem Datensatz T_1 mit 1620 s das bisher schlechteste Ergebnis und ist somit um 1180 s schlechter als bei dem zuvor gezeigten Teilmodell der Kameradaten. Bei dem Datensatz T_2 liegt die Verzugszeit bei einem Wert von 840 s, welche sich um 610 s schlechter als bei dem Teilmodell der Kameradaten und 810 s schlechter als bei dem Modell mit Merkmalsfusion darstellt. Das Ergebnis des Datensatzes T_3 ist problematisch für die Auswertung der Verzugszeit, da die Vorbedingung zu der Definition in Gleichung 7.1.1 nicht erfüllt ist. Problematisch ist, dass das Netzwerk, nach dem Zustandswechsel, noch einmal für einen Zeitschritt die vorherige Klasse berechnet, bevor es konstant bei der gewechselten Klasse bleibt. Wird diese Unsicherheit nicht berücksichtigt, lässt sich eine Verzugszeit von etwa 570 s angeben, welche dann um 110 s länger wäre als bei dem zuvor gezeigten Teilmodell der Kameradaten. Bei der Präzision ist bei den Datensätzen T_2 und T_3 wieder ein Wert von 100 % auszumachen, da bei diesen Datensätzen keine falsch-positiven Klassifikationen auftreten. Bei Datensatz T_1 ist dies jedoch nicht gegeben, sodass sich ein Wert von nur etwa 41 % ergibt. Daraus resultiert für T_1 ein um 31 Prozentpunkte schlechterer Wert und bei dem Datensatz T_3 ein um 17 Prozentpunkte besserer Wert als bei dem Teilmodell der Kameradaten. Bei der Ausfallquote erreicht das hier gezeigte Modell der Herzdaten lediglich bei dem Datensatz T_1 einen Wert von 100 %. Wie zuvor angemerkt, wird bei diesem Bewertungskriterium deutlich, dass der Zustandswechsel für den Datensatz T_3 zu spät stattfindet. Hier liegt im Gegensatz zu dem Modell der Fusion auf Merkmalsebene der Wert der Ausfallquote nicht bei 100 %, sondern etwa 36 Prozentpunkte darunter bei einem Wert von etwa 74 %. Ebenso verhält es sich mit der Ausfallquote des Datensatzes T_2, bei welchem das gezeigte Teilmodell der Herzdaten einen Wert von nur etwa 66 % erreicht. Also schneidet es im Vergleich zu dem vorher gezeigten Teilmodell der Kameradaten 25 Prozentpunkte schlechter ab.

In Abbildung 7.10 sind die Ergebnisse der Testdaten des Teilmodells, welches nur mit Sitzdruckdaten arbeitet, zu sehen. Insgesamt zeigt sich hier ein ähnliches

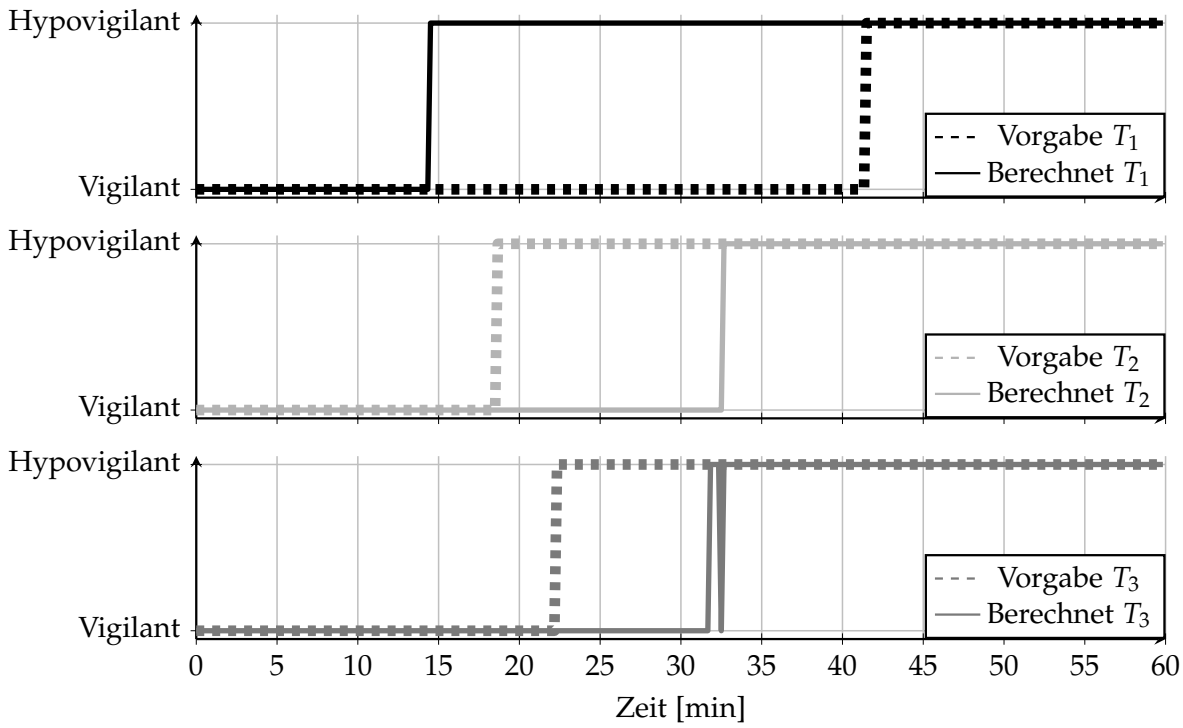


Abbildung 7.9.: Klassifikation der Testdaten von drei Probanden des Hauptversuchs mit dem Teilmodell der Herzdaten

Bild wie bei dem im vorherigen Teilkapitel vorgestellten Modell, welches mit der Fusion auf Merkmalsebene arbeitet, jedoch schneidet es etwas schlechter in diesem Fall ab. Ebenso ist die Berechnung für den Datensatz T_2 am besten, allerdings wird für eine andere Testperson kein Zustandswechsel berechnet. Bei dem Modell der Sitzdruckdaten schlägt die Berechnung für den Datensatz T_3 fehl. Um diesen Eindruck in Zahlen zu belegen, werden die Bewertungskriterien verglichen. Bei der Genauigkeit G ist für dieses Teilmodell ein durchschnittlicher Wert von etwa 70 % anzugeben. Im Einzelnen ergibt sich für den Datensatz T_1 ein Wert von etwa 75 %, welcher gegenüber dem Modell der Merkmalsfusion um etwa ein Prozentpunkt besser ist, da dieses für diesen Datensatz keinen Zustandswechsel berechnet. Für den Datensatz T_2 ist das Ergebnis ähnlich, denn mit nur einem Prozentpunkt Unterschied erreicht das Teilmodell der Sitzdruckdaten eine etwas geringere Genauigkeit G von etwa 98 %. Beim Datensatz T_3 wird von dem Modell der Sitzdruckdaten kein Zustandswechsel berechnet. Da bei diesem Datensatz der Zustandswechsel schon etwa in Minute 22 stattfindet und das Modell der Sitzdruckdaten über die komplette Zeit die Klasse Vigilant berechnet, liegt der Wert für die Genauigkeit G nur bei 37 %. Damit schneidet es bisher am schlechtesten bei diesem Datensatz ab, denn es ist um etwa 47 Prozentpunkte schlechter als das Modell der Fusion auf Merkmalsebene. Bei der Verzugszeit t_v ergibt sich für das Modell der Sitzdruckdaten bei dem Datensatz T_1 ein Wert von 890 s, welcher nicht mit dem Modell der Fusion auf Merkmalsebene verglichen werden kann, da kein Zustandswechsel berechnet wurde. Im Vergleich zu dem Teilmodell der Kameradaten, ist das zuletzt gezeigte Modell um 450 s schlechter und zu dem Teilmodell der Herzdaten um 730 s besser. Bei dem Datensatz

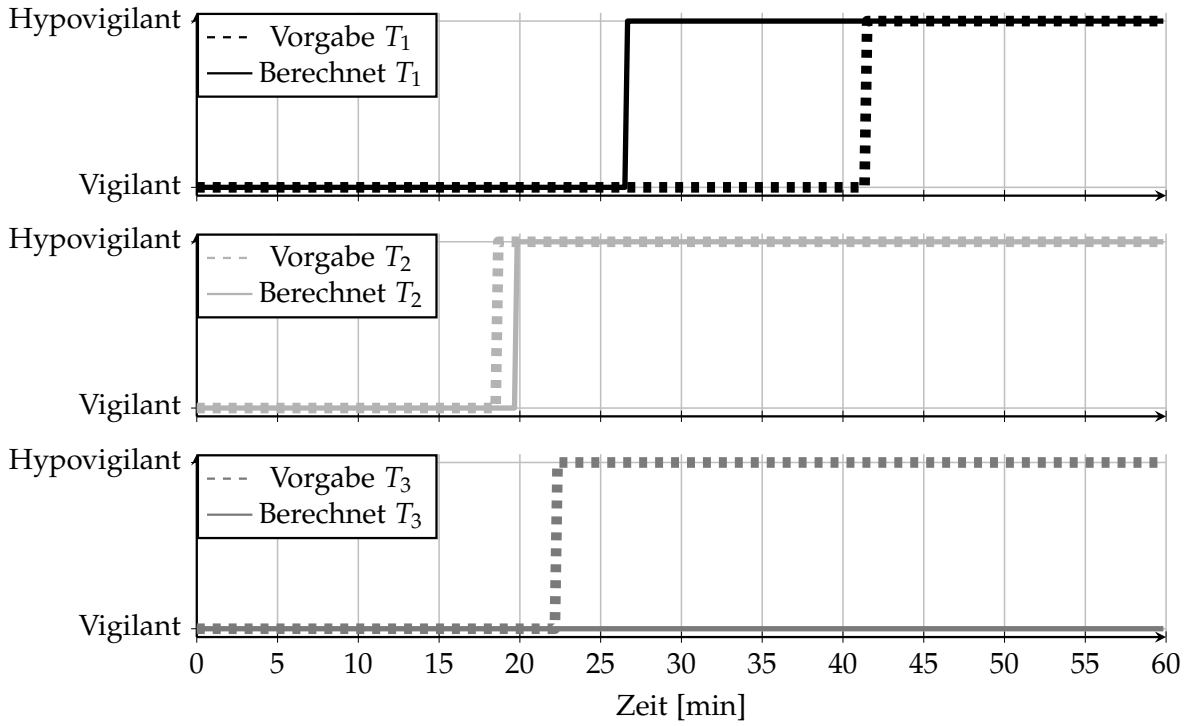


Abbildung 7.10.: Klassifikation der Testdaten von drei Probanden des Hauptversuchs mit dem Teilmodell der Sitzdruckdaten

T_2 ergibt sich für das Teilmodell der Sitzdruckdaten eine um 40 s schlechtere Verzugszeit als bei dem Modell der Fusion auf Merkmalsebene, denn es wird eine Verzugszeit von 70 s erreicht. Für den Datensatz T_3 lässt sich von dem Teilmodell der Sitzdruckdaten keine Verzugszeit angeben, da von diesem kein Zustandswechsel berechnet wird. Dies wirkt sich auch auf den Präzisionswert P aus, der für diesen Datensatz ebenfalls nicht angegeben werden kann, da weder falsch-positive noch richtig-positive Klassifikationen auftreten. Für den Datensatz T_2 ergibt sich mit 100 % das gleiche Ergebnis wie bei dem Modell der Klassifikation auf Merkmalsebene. Da bei dem Datensatz T_1 für das zuletzt genannte Modell keine Angabe zum Wert der Präzision möglich ist, wird dieser in Bezug zu den anderen Teilmodellen gesetzt. Hierbei erreicht das Teilmodell der Sitzdruckdaten einen Wert von etwa 56 % und ist damit um 16 Prozentpunkte schlechter als das Teilmodell der Kameradaten sowie 16 Prozentpunkte besser als das Teilmodell der Herzdaten. Bei der Ausfallquote ergeben sich für die Datensätze T_1 und T_3 gegensätzliche Ergebnisse zum Modell der Fusion auf Merkmalsebene, denn bei dem einen ist der Wert aufgrund des ausbleibenden Wechsels Null und bei dem jeweils anderen bei 100 % da keine falsch-negativen Klassifikationen auftreten. Bei Datensatz T_2 wird ein Wert von etwa 98 % für die Ausfallquote R erreicht, welcher nur um einen Prozentpunkt schlechter ist als bei dem Modell der Fusion auf Merkmalsebene.

Nachdem die Performanz der einzelnen Teilmodelle besprochen wurde, erfolgt nun die Auswertung der einzelnen Klassenwahrscheinlichkeiten der Teilmodelle für jeden der Testdatensätze. Die Auswertung für den Datensatz T_1 ist in Abbildung 7.11 dargestellt. Die Klassenwahrscheinlichkeiten werden für nur eine der beiden Klassen

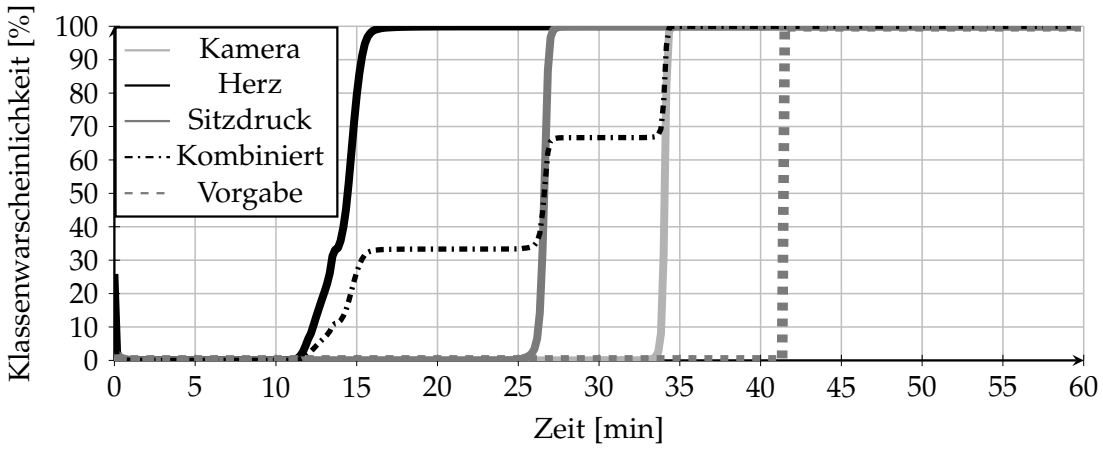


Abbildung 7.11.: Kombination der Klassenwahrscheinlichkeiten für den Datensatz $T1$

dargestellt, da die Klassenwahrscheinlichkeit für die andere Klasse hierzu, wie zuvor schon beschrieben, komplementär ist. Zu den drei Klassenwahrscheinlichkeiten der einzelnen Teilmodelle (durch durchgezogene Linien dargestellt) wird die kombinierte Klassenwahrscheinlichkeit und die Vorgabe, gegeben durch die Fragebögen, über die Versuchszeit von 60 Minuten dargestellt. Hierbei wird deutlich, dass die einzelnen Teilmodelle jeweils einen schnellen Wechsel von einer Klassenwahrscheinlichkeit von 0 % auf eine Klassenwahrscheinlichkeit von 100 % aufweisen. Dieser Wechsel läuft bei dem Teilmodell der Kamera- und Sitzdruckdaten in unter einer Minute ab und bei dem Teilmodell auf Basis der Herzdaten in etwa vier Minuten. Nach dem Wechsel bleiben alle Teilmodelle sehr zuversichtlich ohne eine weitere Änderung der Klassenwahrscheinlichkeit bei der entsprechenden Klassifikation. So ergibt sich für die kombinierte Klassenwahrscheinlichkeit eine Art Stufenform, da der Wert bei jedem Wechsel der Klassenwahrscheinlichkeit um ein Drittel zunimmt. Damit ist das Ergebnis der Klassifizierung nahezu identisch zu der Klassifikation des Teilmodells der Sitzdruckdaten, da die anderen Modelle kaum eine Änderung in diesem Bereich aufweisen. Der Zustandswechsel wird an dem Zeitpunkt vom Modell der kombinierten Klassenwahrscheinlichkeiten berechnet, an dem zwei der drei Teilmodelle die entsprechende Klasse auswählen. So entsteht ein Mehrheitsentscheid, welche Klasse von dem Modell ausgegeben wird. Damit erreicht das kombinierte Modell eine um etwa 2 Prozentpunkte höhere Genauigkeit G gegenüber dem Mittelwert der drei einzelnen Teilmodelle für diesen Datensatz T_1. Aus Tabelle 7.8 lassen sich die weiteren Auswertekriterien entnehmen.
Für den zweiten Datensatz T_2 ist die Auswertung der Klassenwahrscheinlichkeiten der verschiedenen Teilmodelle über die Zeit in Abbildung 7.12 abgebildet. Bei diesem Datensatz ist ein Unterschied sehr deutlich, denn das Teilmodell der Herzdaten zeigt ein differenziertes Verhalten. Die Klassenwahrscheinlichkeit steigt nicht, wie bei den anderen beiden Teilmodellen, schnell in kurzer Zeit an, sondern über längere Zeit mit einer gewissen Varianz. Dieser Prozess findet etwa von Minute 14 bis 38 statt. Danach bleibt auch die Klassenwahrscheinlichkeit von diesem Modell auf einem gleichbleibend hohen Niveau. Die anderen zwei Teilmodelle haben eine sehr kurze Anstiegszeit von unter einer Minute. Vor und nach diesem schnellen Wechsel sind

Tabelle 7.8.: Konfusionsmatrix Modell mit Fusion auf Entscheidungsebene mit dem Testdatensatz $T1$

		Berechnet		
		Hypovigilant	Vigilant	
Vorgabe	Hypovigilant	111	0	$R = 100,00\,\%$
	Vigilant	89	160	$S = 64,26\,\%$
		$P = 55,50\,\%$	$V = 100,00\,\%$	$G = 75,28\,\%$

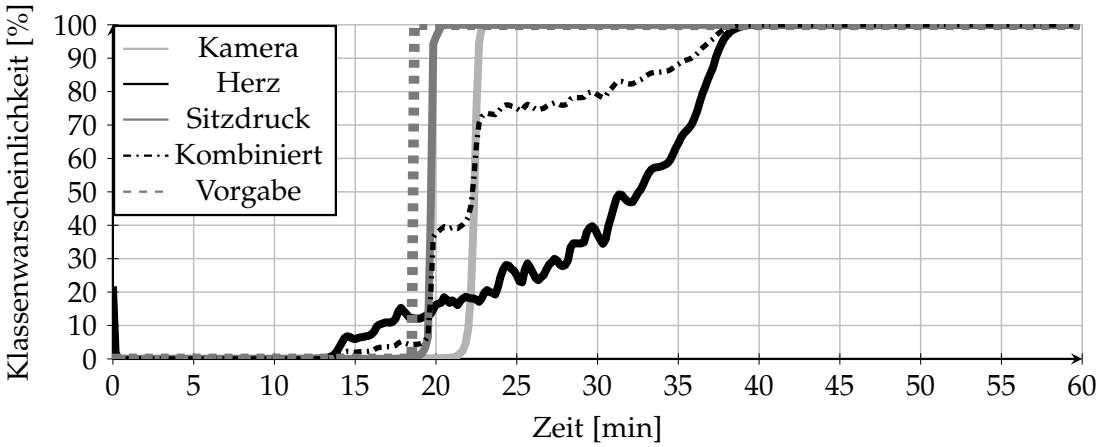


Abbildung 7.12.: Kombination der Klassenwahrscheinlichkeiten für den Datensatz $T2$

bei beiden Teilmodellen die Klassenwahrscheinlichkeiten stetig bei etwa 0 % oder 100 %. Dazu wird, wie bei den Daten aus dem Vorversuch, in den ersten Minuten eine kleinere Unsicherheit deutlich, welche bei allen Teilmodellen auftritt. Somit ist die kombinierte Klassenwahrscheinlichkeit wie Stufen geformt, jedoch ebenso geprägt von dem langsamen Anstieg des Teilmodells der Herzdaten. Letzteres führt dazu, dass der Zustandswechsel von dem auf Entscheidungsebene kombinierten Modell minimal früher geschieht als bei dem Teilmodell der Kameradaten. Dieser Effekt ist so gering, dass sich für diese Auswertung nur ein Unterschied von einem Zeitschritt, also 10 s, ergibt. Dennoch ist das kombinierte Modell in der Genauigkeit besser als zwei der Teilmodelle. So ergibt sich ebenfalls eine Verbesserung der Genauigkeit gegenüber der durchschnittlichen Genauigkeit der Teilmodelle mit einem Unterschied von etwa fünf Prozentpunkten. Die durchschnittliche Genauigkeit der Teilmodelle für den Datensatz T_2 liegt bei etwa 89 % und die des kombinierten Modells bei etwa 94 %. Die anderen Auswertekriterien lassen sich in Tabelle 7.9 finden.

Das Modell auf Basis der Fusion auf Entscheidungsebene wird dazu noch mit dem dritten Testdatensatz T_3 ausgewertet. Die Klassenwahrscheinlichkeiten zu den einzelnen Teilmodellen und dem kombinierten Verfahren sowie eine Einordnung der Vorgabe wird in Abbildung 7.13 gezeigt. Wie schon vorher deutlich wurde, ist bei diesem Datensatz das Modell der Sitzdruckdaten von

Tabelle 7.9.: Konfusionsmatrix Modell mit Fusion auf Entscheidungsebene mit dem Testdatensatz $T2$

		Berechnet		
		Hypovigilant	Vigilant	
Vorgabe	Hypovigilant	226	22	$R = 91,13\,\%$
	Vigilant	0	112	$S = 100,00\,\%$
		$P = 100,00\,\%$	$V = 83,58\,\%$	$G = 93,89\,\%$

allen Teilmodellen am schlechtesten, denn in der Klassenwahrscheinlichkeit wird keine Änderung deutlich. Über die komplette Versuchszeit bleibt die Klassenwahrscheinlichkeit für die sogenannte Klasse Hypovigilant bei null Prozent. Die Klassenwahrscheinlichkeiten der anderen beiden Teilmodelle verhalten sich analog zu den anderen beiden Testdatensätzen. Auch bei dem Testdatensatz T_3 wird deutlich, dass bei dem Teilmodell der Kameradaten der Wechsel zwischen kleinen und großen Klassenwahrscheinlichkeitswerten wieder in unter 60 s verläuft. Dahingegen ist bei der Klassenwahrscheinlichkeit des Teilmodells der Herzdaten wieder das zuvor beobachtete Verhalten zu sehen. Die Klassenwahrscheinlichkeit steigt ab etwa Minute 13 langsam von null Prozent mit einer gewissen Varianz bis etwa Minute 38 nur langsam an. Bei der kombinierten Klassenwahrscheinlichkeit ergibt sich somit wieder ein sprunghafter Anstieg etwa in Minute 14 und danach dann ein langsamer Anstieg analog zu dem Verhalten des Teilmodells der Herzdaten. Wieder gilt, wenn zwei der drei Modelle den Zustandswechsel berechnen, wird auch das Modell mit der Fusion auf Entscheidungsebene den entsprechenden Zustandswechsel berechnen. Das Teilmodell auf Basis der Sitzdruckdaten weist eine sehr schlechte Genauigkeit von etwa 37 % auf, sodass die Genauigkeit von dem Modell der Fusion auf Entscheidungsebene besser abschneidet als die durchschnittliche Genauigkeit der drei Teilmodelle. Dabei erreicht das Modell auf Basis der Fusion auf Entscheidungsebene einen Wert der Genauigkeit von etwa 84 % und ist somit um 15 Prozentpunkte besser als der durchschnittliche Genauigkeitswert der Teilmodelle, der bei einem Wert von etwa 69 % liegt. Weitere Auswertekriterien lassen sich der Tabelle 7.10 entnehmen.

Untersuchung der Ausfallsicherheit

Bei den Studien ist aufgefallen, dass bei einigen Messungen die Zuverlässigkeit der Sensoren nicht immer gegeben ist. So wurden bei dem Hauptversuch fünf Probanden aus der Trainingsmenge entnommen, da die Daten nur fehlerhaft vorliegen. Für das Training ist es wichtig, dass die Daten zu jedem Zeitpunkt vorliegen, sodass die Modelle die Möglichkeit haben, sich einzustellen. Für eine realistische Einschätzung der Performanz ist es angebracht, auch unvollständige Datensätze zu berücksichtigen. So könnte ein Abbruch der Signalqualität durch

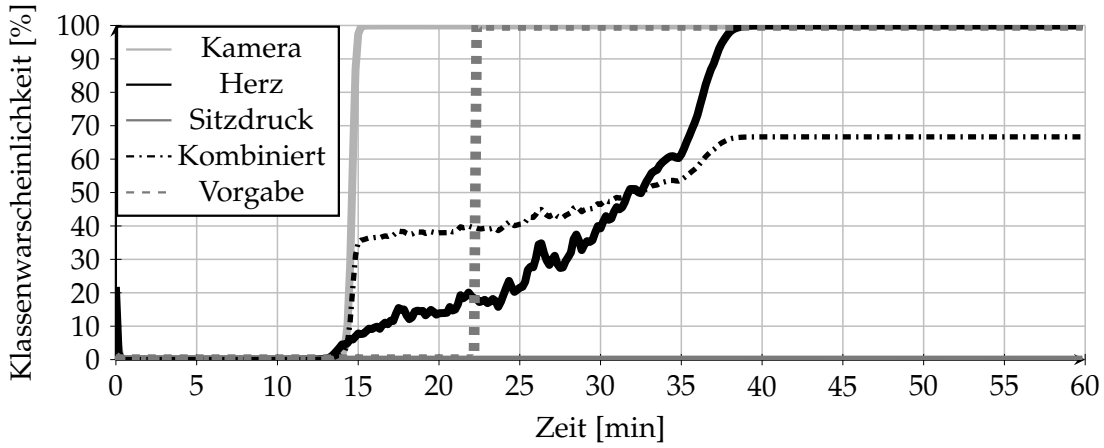


Abbildung 7.13.: Kombination der Klassenwahrscheinlichkeiten für den Datensatz $T3$

Tabelle 7.10.: Konfusionsmatrix Modell mit Fusion auf Entscheidungsebene mit dem Testdatensatz $T3$

		Berechnet		
		Hypovigilant	Vigilant	
Vorgabe	Hypovigilant	168	58	$R = 74,34\,\%$
	Vigilant	0	134	$S = 100,00\,\%$
		$P = 100,00\,\%$	$V = 69,79\,\%$	$G = 83,89\,\%$

starke Bewegung oder Umgebungseffekte auftreten. Um die Performanz auch unter diesen Umständen zu bewerten, wird zunächst einer der fünf aussortierten Probanden-Datensätze genutzt. In Abbildung 7.14 sind die Merkmale 1-4 aus Tabelle 6.2 für diesen fehlerhaften Datensatz ($F1$) dargestellt. Auf der x-Achse ist bei allen vier Teilabbildungen die Versuchszeit von 60 Minuten dargestellt. Auf der y-Achse ist in jeder Teilabbildung der Wert für das normalisierte und vom Mittelwert befreite Merkmal abgebildet. Durch die Normalisierung werden die Daten einheitslos und bewegen sich um null herum mit einer Varianz, die in den meisten Fällen kleiner als zwei beträgt. In den beiden oberen Teilabbildungen sind die Merkmale des Herzsensors zu sehen. Dieser hat über die komplette Versuchszeit plausible Daten hervorgebracht. Dabei sei hervorzuheben, dass die y-Achse bei der zweiten Teilabbildung eine stark abweichende Skalierung im Gegensatz zu den anderen Merkmalsabbildungen aufweist. Hier bewegt sich das Signal nur minimal um einen Wert von etwa -0,175. Diese Normierung kann dadurch zustande gekommen sein, dass in den Daten öfter große Ausreißer vorliegen. Diese könnten dazu führen, dass die Normierung bei diesem Signal entsprechend stark ausfällt und in den meisten Fällen sehr klein skaliert. Im Rahmen der vorliegenden Arbeit wurde keine tiefere Analyse der Rohdaten durchgeführt, um Gegenmaßnahmen zu diesem Verhalten zu treffen. Denkbar wäre es, die Rohdaten vor der weiteren Verarbeitung zu filtern, um unerwünschte Artefakte aus den Daten zu entfernen. Dies könnte durch ein gewisses Toleranzband geschehen, ähnlich zu der Berechnung der Augenschließdauer. In den beiden unteren Teilabbildungen sind die Merkmale dargestellt, die von dem Kamerasensor erfasst werden. Bei den Abbildungen wird deutlich, dass bei diesem Datensatz eine Lücke in den Daten entstanden ist. Zwischen etwa Minute sechs bis etwa Minute 56 liegen keine Daten vor. In diesem Fall wurde für den entsprechenden Zeitraum für jeden Zeitschritt ein Wert von null angenommen. Da die Daten vom Mittelwert befreit sind, sollten sie so am neutralsten für die Modelle sein. Für die Berechnung der Klassifizierung ist es notwendig, dass für jeden Zeitschritt ein Wert für die entsprechenden Merkmale vorliegt. Damit die Ergebnisse für die beiden Fusionsmethoden so vergleichbar wie möglich bleiben, wird auch bei dem Modell mit Fusion auf Entscheidungsebene ein angepasster Datensatz verwendet. Denkbar wäre, für das gerade genannte Modell den Datensatz nicht anzupassen und das Teilmodell der Kameradaten nicht zu verwenden. Ebenso könnte für die Zeit der fehlenden Daten eine andere Methode angewandt werden. So könnte das Wissen über die Form der Daten genutzt werden, um eine Art Simulation für die Zeit der fehlenden Daten zu erzeugen. Eine tiefere Analyse dieser Methoden hat im Rahmen der Arbeit nicht stattgefunden. Die Daten der Sitzdrucksensoren sind aufgrund der Menge nicht dargestellt. Zudem kann bei dem Sensor kein Datensatz ermittelt werden, bei dem eine Lücke bei diesen Merkmalen aufgetreten ist.

Dieser angepasste Datensatz soll nun mit den beiden Fusionsmethoden getestet werden. Die Erwartung ist dabei, dass die Fusion auf Merkmalsebene schlechter abschneiden wird als die Fusion auf Entscheidungsebene. Wie zu Anfang der vorliegenden Arbeit in der Beschreibung der Literatur schon beschrieben wurde, sollten getrennte künstliche Netzwerke eine größere Ausfallsicherheit mit sich bringen.

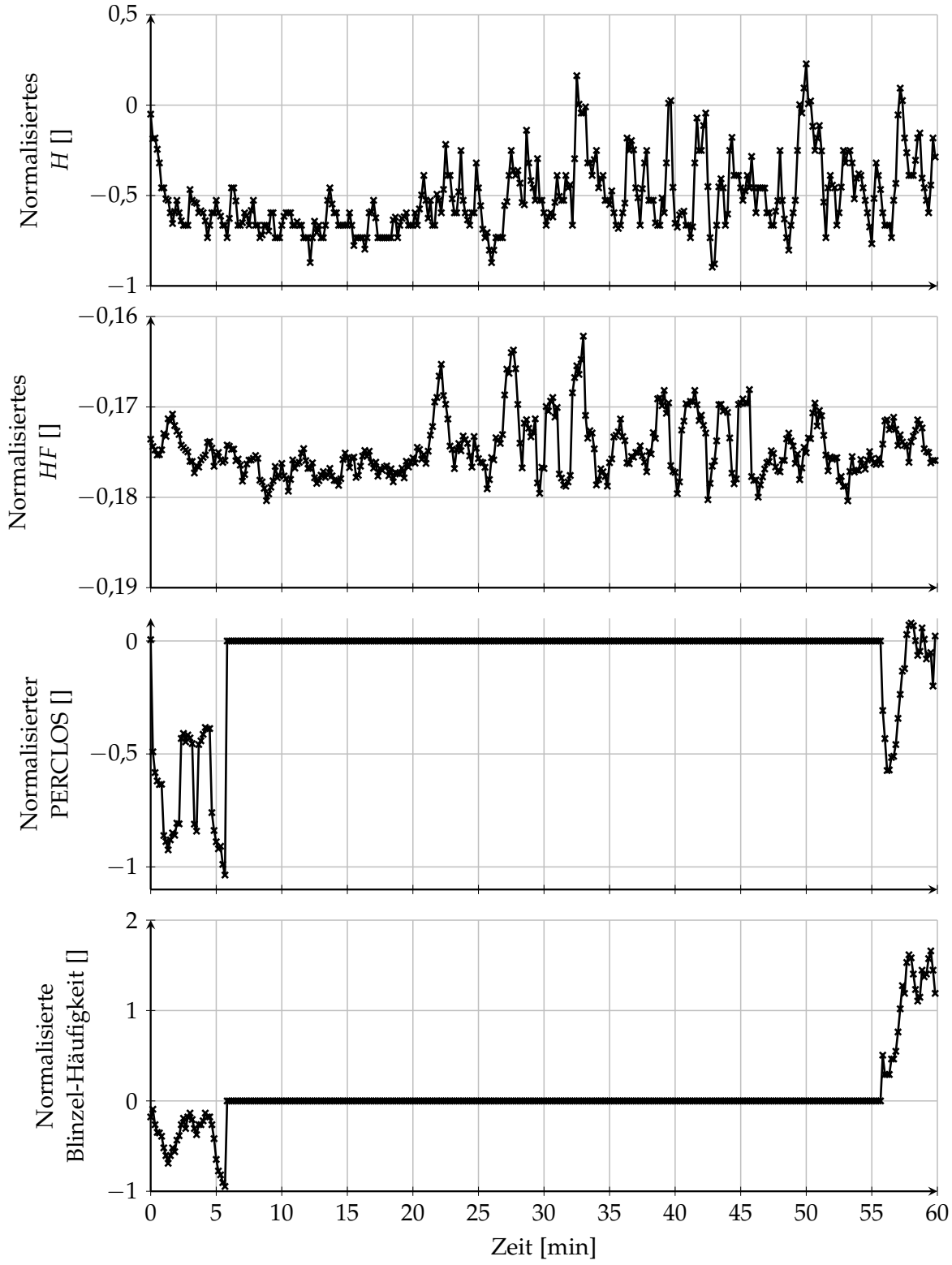


Abbildung 7.14.: Merkmalsdaten für einen Probanden mit Datenlücke

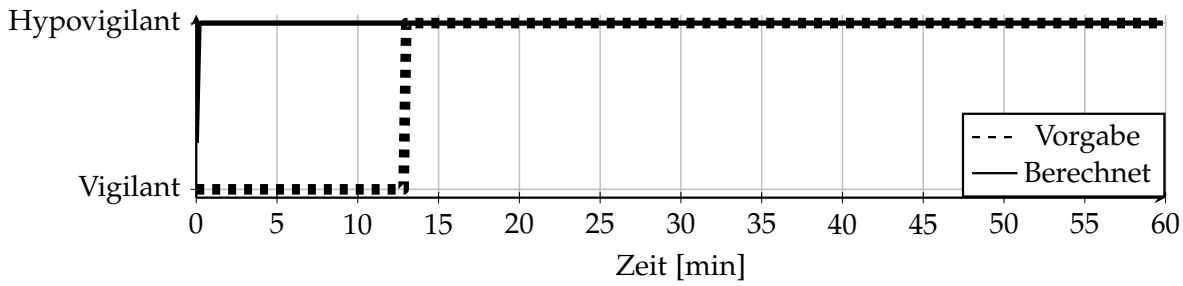


Abbildung 7.15.: Klassifikation des Datensatzes F_1 mit allen verfügbaren Merkmalen und der Fusion auf Merkmalsebene

Das Ergebnis des Modells der Fusion auf Merkmalsebene ist in Abbildung 7.15 dargestellt. Dabei ist, wie in den Abbildungen zuvor, auf der x-Achse die Versuchszeit von 0 bis 60 Minuten und auf der y-Achse das Ergebnis des Modells sowie die Vorgabe aus den Probanden-Fragebögen zu finden. In dieser Abbildung wird deutlich, dass nur ein einziger Zustandswechsel auftritt. Dies geschieht im allerersten Zeitschritt. Danach bleibt das Ergebnis der Klassifikation bei der gleichen Klasse. Somit ergibt sich bei dem Datensatz F_1 eine Genauigkeit G von lediglich 78,61 %. Dieses Ergebnis erzielt nur so einen hohen Wert, da der Zustandswechsel bei diesem Datensatz schon recht früh in Minute 13 aus dem Probanden-Fragebogen hervorgeht und das Netzwerk ohne einen Zustandswechsel schon bei einer Genauigkeit von 78,33 % liegen würde. Die Verzugszeit t_v liegt bei einem Wert von 770 s, welche im Vergleich zu den vorher gezeigten Modellen vergleichsweise lang ist. Die Präzision P ist bei diesem Datensatz F_1 bei einem Wert von 78,55 %, wohingegen der Wert für die Ausfallquote bei 100 % liegt. Diese Werte ergeben sich durch die hohe Anzahl von 77 falsch-positiven Klassifikationsergebnissen und dem Ausbleiben von falsch-negativen Klassifikationen.

In Abbildung 7.16 sind die Klassenwahrscheinlichkeiten für den Datensatz F_1 für die einzelnen Teilmodelle und kombinierten Verfahren dargestellt. Wie zuvor wird auf der x-Achse die Versuchszeit von 60 Minuten und auf der y-Achse die Klassenwahrscheinlichkeit für die hier sogenannte Klasse Hypovigilant aufgetragen. Hierbei kann festgestellt werden, dass die Klassenwahrscheinlichkeit von dem Teilmodell auf Basis der Kameradaten, welche bei diesem Datensatz lückenhaft sind, das beste Klassifikationsergebnis zeigt. Die Klassenwahrscheinlichkeit verhält sich, so wie zuvor beobachtet, mit einem sehr schnellen Anstieg in kurzer Zeit in etwa Minute 17. Vor und nach diesem schnellen Wechsel bleibt sie recht konstant bei einem Wert von null beziehungsweise 100 %. Der Verlauf der Klassenwahrscheinlichkeit des Teilmodells auf Basis der Herzdaten ist ebenso vergleichbar mit den zuvor gezeigten Verläufen. In diesem Fall zieht sich der Anstieg der Klassenwahrscheinlichkeit über mehr als 10 Minuten, von etwa Minute 13 bis etwa Minute 25. Bei dem Teilmodell auf Basis der Sitzdruckdaten hingegen bleibt die Klassenwahrscheinlichkeit bei einem Wert von 100 % über die komplette Versuchszeit. Somit ergibt sich der Verlauf der Klassenwahrscheinlichkeit des kombinierten Modells. Da das Teilmodell auf Basis der Sitzdruckdaten von Beginn an bei einer Klassenwahrscheinlichkeit von 100 % liegt, startet das kombinierte Modell ebenso bei einem Wert von etwa 33 % und

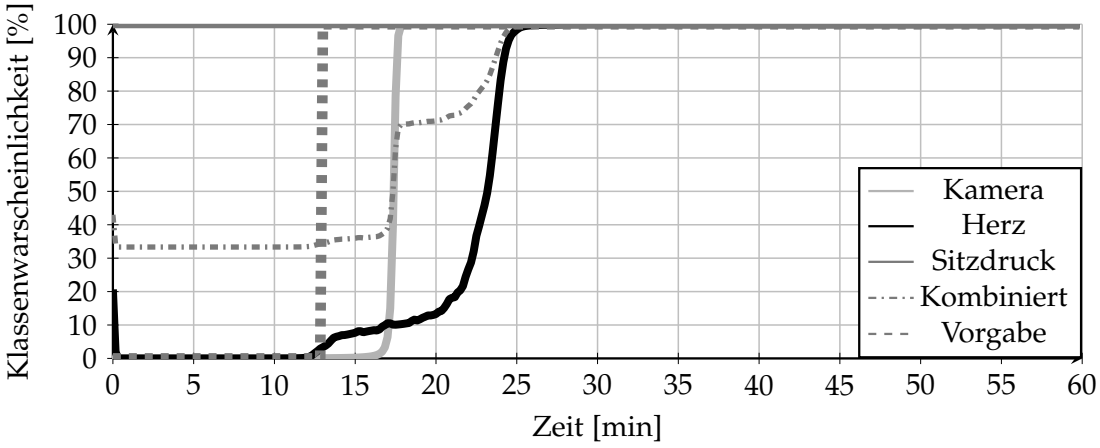


Abbildung 7.16.: Kombination der Klassenwahrscheinlichkeiten für den Datensatz F_1

unterschreitet diesen nicht. Mit dem schnellen Zustandswechsel des Teilmodells auf Basis der Kameradaten überschreitet auch das kombinierte Modell die für den Zustandswechsel relevanten 50 % der Klassenwahrscheinlichkeit in Minute 17. Damit erreicht das Modell auf Basis der kombinierten Klassenwahrscheinlichkeiten eine Genauigkeit G von 92,5 % und ist somit um acht Prozentpunkte besser als der Mittelwert der Genauigkeit der drei einzelnen Teilmodelle. Verglichen mit der zuvor gezeigten Genauigkeit G des Modells, welches auf Fusion in der Merkmalsebene basiert, gibt es einen Unterschied von etwa 14 Prozentpunkten. Hierbei wird die zuvor beschriebene These gestützt, da das Modell auf Basis der Fusion auf Entscheidungsebene tatsächlich für den Datensatz F_1 eine bessere Performanz zeigt als das Modell auf Basis der Fusion auf Merkmalsebene.
Da ein Datensatz alleine jedoch wenig Aussagekraft hat, wird ein zweiter Datensatz, hier F_2 genannt, mit dem gleichen Hintergrund untersucht. Dieser ist nicht in jeder Hinsicht vergleichbar, da die Sensoren nicht zur gleichen Zeit ausfallen. In diesem Fall sind auch nicht die Kameradaten, sondern die Daten des Herzsensors, betroffen. Die Daten sind an mehreren Zeitpunkten entweder nicht plausibel oder erst gar nicht vorhanden. Da eine Korrektur der Merkmale deutlich aufwändiger wäre, wird dieser Datensatz, so wie er aufgezeichnet wurde, für die Auswertung verwendet.
In Abbildung 7.17 ist das Ergebnis der Auswertung des Datensatzes F_2 mit dem Modell auf Basis der Fusion auf Merkmalsebene zu sehen. Bei dieser ist wie zuvor auf der x-Achse die Versuchszeit und auf der y-Achse das Ergebnis der Klassifikation abgebildet. Auch bei diesem Datensatz wird deutlich, dass das Modell keine sinnvolle Reaktion auf die Daten zeigt. Es wird also kein Zustandswechsel von dem Modell berechnet. Somit ergibt sich eine Genauigkeit G von 62,22 %, welche somit um etwa 16 Prozentpunkte schlechter ausfällt als in dem zuvor gezeigten Beispiel mit dem Datensatz F_1. Dieser Unterschied ist auf die unterschiedlichen Vorgaben aus den Probanden-Fragebögen zu erklären. In dem Fall kann, wie zuvor schon beschrieben, keine Verzugszeit t_v angegeben werden, da kein Zustandswechsel von dem Modell berechnet wird. Somit ist ebenfalls die Angabe eines Präzisionswertes P für diese Auswertung nicht möglich, da es weder falsch- noch richtig-positiven Klassifikationen gibt. Für die Ausfallquote R ergibt sich ein Wert von 0 %, welcher ebenfalls durch die

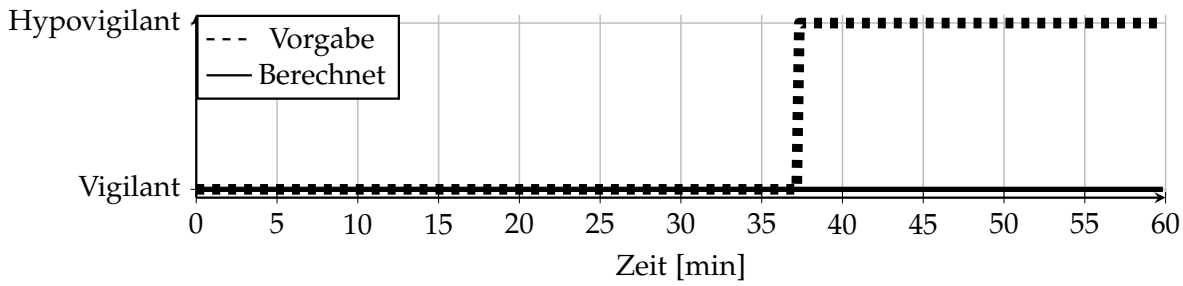


Abbildung 7.17.: Klassifikation des Datensatzes F_2 mit allen verfügbaren Merkmalen und der Fusion auf Merkmalsebene

ausbleibenden richtig-positiven Klassifikationen zustandekommt.
In der Abbildung 7.18 sind die Klassenwahrscheinlichkeiten der Teilmodelle sowie die kombinierte Klassenwahrscheinlichkeit für das Modell auf Basis der Fusion auf Entscheidungsebene für den Datensatz F_2 dargestellt. Hierbei sind ebenfalls auf der x-Achse die Versuchszeit und auf der y-Achse die Klassenwahrscheinlichkeit für die hier sogenannte Klasse Hypovigilant abgebildet. Die Klassenwahrscheinlichkeit des Teilmodells auf Basis der Kameradaten verhält sich wie zuvor beobachtet. Auch bei diesem Datensatz F_2 wird ein schneller Wechsel innerhalb weniger Sekunden berechnet. Dabei ist eine kleine Unsicherheit bei etwa Minute 26 am Anfang des schnellen Wechsels zu erkennen. Vor und nach dem Zustandswechsel bleibt die Klassenwahrscheinlichkeit bei null beziehungsweise 100 %. Bei dem Teilmodell auf Basis der Kameradaten, die bei diesem Testdatensatz F_2 fehlerhaft sind, wird dennoch ein Zustandswechsel berechnet. In den ersten Minuten zeigt sich auch hier eine Unsicherheit in der Klassenwahrscheinlichkeit. Bis zu dem schnell verlaufenden Zustandswechsel in etwa Minute acht bleibt die Klassenwahrscheinlichkeit konstant um null Prozent. Nach dem Zustandswechsel bleibt diese ebenso kontinuierlich bei einem Wert um 100 %. Bei anderen Datensätzen konnte bei diesem Modell auch schon ein langsamer Anstieg der Klassenwahrscheinlichkeit beobachtet werden, jedoch war dieser bei dem Testdatensatz T_1 ähnlich schnell. Das Teilmodell der Sitzdruckdaten zeigt keine Reaktion auf den Datensatz und bleibt konstant bei einer Klassenwahrscheinlichkeit von null Prozent. Damit verläuft die kombinierte Klassenwahrscheinlichkeit, wie auch zuvor beobachtet, in Stufen. Dabei wird die erste Stufe etwa in Minute acht durch den Zustandswechsel des Teilmodells auf Basis der Herzdaten von etwa null Prozent auf etwa 33 Prozent hervorgerufen. Die zweite Stufe wird dann etwa in Minute 26 berechnet, da hier der Zustandswechsel des Teilmodells auf Basis der Kameradaten stattfindet. In diesem Fall steigt die Klassenwahrscheinlichkeit von dem Wert 33 % etwa auf den Wert 66 %, sodass auch der Zustandswechsel des kombinierten Modells berechnet wird, da der Wert von 50 Prozentpunkten überschritten wird. Damit ergibt sich eine Genauigkeit G für den Datensatz F_2 bei der Fusion auf Entscheidungsebene von etwa 81,67 %. Der genannte Wert ist ebenso gegenüber dem Mittelwert der einzelnen Teilmodelle besser, denn dieser liegt bei 64,91 % und ist somit um etwa 17 Prozentpunkte schlechter. Auch bei Datensatz F_2 wird die oben genannte These gestützt, dass eine Fusion auf Entscheidungsebene eine größere Ausfallsicherheit mit sich bringt als die Fusion auf

Merkmalsebene.

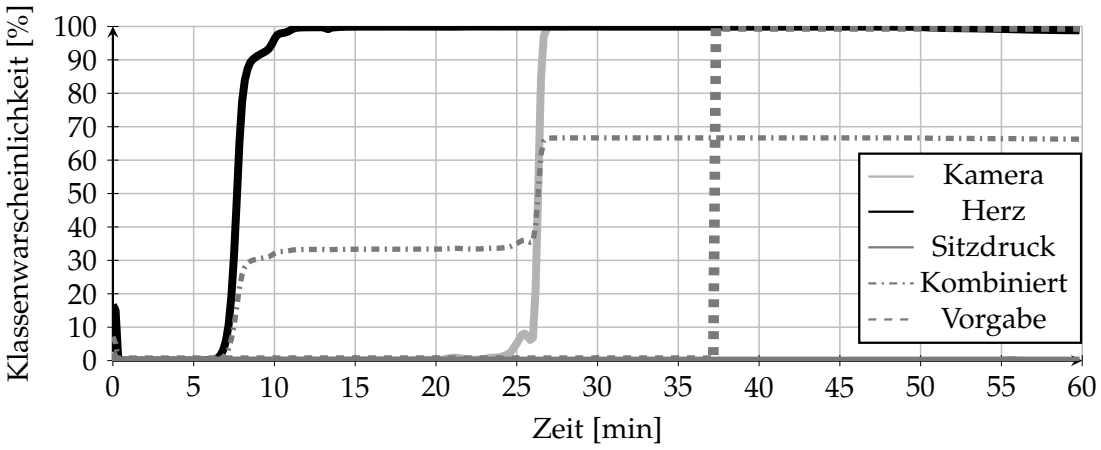


Abbildung 7.18.: Kombination der Klassenwahrscheinlichkeiten des Datensatzes F_2

7.2. Modell zur Bestimmung des Aufmerksamkeitsverhaltens

Nachdem nun die Ergebnisse im Detail dargestellt wurden, soll eine Bewertung der beiden Modelle erfolgen. Zum einen wird in der vorliegenden Arbeit das Modell auf Basis der Fusion auf Merkmalsebene und zum anderen das Modell mit Fusion auf Entscheidungsebene vorgestellt. Beide Ansätze werden mit den Daten aus den beiden Versuchen, der Vor- und der Hauptstudie, trainiert und getestet. Da exakt dieselben Datensätze für beide Modelle zur Verfügung standen, sollte eine Vergleichbarkeit vorhanden sein. Nachfolgend wird die Bewertung in zwei Teile aufgeteilt, da sowohl das Training als auch die Performanz der Modelle ausschlaggebend sind.

Training

Die Dauer des Trainings ist ein Faktor bei der Erstellung von Modellen auf Basis von künstlichen neuronalen Netzwerken. Da das Training nicht im Fokus steht, wird lediglich eine qualitative Einschätzung zu dieser Thematik gegeben. Die absolute Dauer des Trainings hängt von verschiedenen Faktoren ab, wie zum Beispiel der verwendeten Hardware, dem vorhandenen Datensatz sowie der Größe des Modells und damit auch der zu trainierenden Variablen. Einige dieser Faktoren bleiben in der hier vorliegenden Arbeit für beide Modellansätze gleich, wie die Hardware und die Datensätze. Die Menge der zu trainierenden Variablen unterscheidet sich jedoch bei den beiden Modellen. So wird für jedes Teilmodell bei der Fusion auf Entscheidungsebene ein eigenes Klassifikationsnetzwerk verwendet, anstatt einem Klassifikationsnetzwerk bei dem Modell auf Basis der Merkmalsfusion. So wird bei dem Modell mit der Fusion auf Merkmalsebene nur ein aufwendiger Trainingsprozess benötigt, wohingegen bei der Fusion auf Entscheidungsebene drei recht äquivalente Trainingsprozesse durchgeführt werden müssen. Da sich für jedes der drei Teilmodelle die Größe der Datenmenge am Eingang des Netzwerks unterscheidet, dauert das

Training allerdings nicht dreimal so lange, wie bei dem einen Modell bei der Fusion auf Merkmalsebene, aber dennoch deutlich länger. Dazu muss noch berücksichtigt werden, dass die zu trainierenden Variablen am Anfang zufällig initialisiert werden und somit das Modell nur ein lokales Minimum innerhalb eines Trainingsprozesses erreichen könnte. Aus diesem Grund wurde das Training, in der hier vorliegenden Arbeit, mehrmals wiederholt, um ein globales Minimum zu finden. Dabei wurde der Trainingsprozess mindestens 10 Mal wiederholt. So wirkt sich der gerade beschriebene Unterschied nochmal mehr aus. Damit ist im Hinblick auf das Training das Modell auf Basis der Merkmalsfusion besser geeignet, da für ein erneutes Training bei verändertem Datensatz weniger Zeit in Anspruch genommen werden muss.

Performanz

Die Performanz der beiden Modellansätze wurde im Kapitel 7.1 ausführlich dargestellt, jedoch nicht direkt gegenübergestellt. Dies soll nun im Folgenden geschehen, indem die Performanz der verschiedenen Modelle mit den unterschiedlichen Datensätzen im Mittel verglichen wird. Dabei wird nur eines der Bewertungskriterien herangezogen. In diesem Fall ist die Entscheidung auf die Genauigkeit G gefallen, da dieses Kriterium in jedem Fall berechnet werden kann und somit in jede der Mittelwertbildungen einfließt. Die Gegenüberstellung kann für drei verschiedene Testdatensätze erfolgen. Hierbei kann zunächst der Testdatensatz aus der Vorstudie verwendet werden. Darüber hinaus steht der Testdatensatz der Hauptstudie zur Verfügung und zudem noch die Auswertung der unvollständigen Testdatensätzen, folgend Ausfallsicherheit genannt. Die Ergebnisse zu den einzelnen Datensätzen sind in Abbildung 7.19 dargestellt. Für die Ergebnisse der Vorstudie ist keine Mittelwertberechnung notwendig, da der Testdatensatz nur aus den Daten einer Testperson besteht. Für die Ergebnisse der Hauptstudie wird der Mittelwert über die Genauigkeit der drei Testpersonen T_1 bis T_3 berechnet. Die hier dargestellten Ergebnisse der Ausfallsicherheit ergeben sich aus den Mittelwerten der Genauigkeiten der zwei Testpersonen F_1 und F_2.
Bei den Ergebnissen der Vorstudie ergibt sich ein Unterschied zwischen den beiden Modellen. Das Modell auf Basis der Fusion auf Merkmalsebene erzielt eine um fünf Prozentpunkte höhere durchschnittliche Genauigkeit, mit einem Wert von 96 %, als das Modell auf Basis der Fusion auf Entscheidungsebene, welches 91 % erreicht. Somit kann die Einschätzung getroffen werden, dass sich das Modell auf Basis der Fusion auf Merkmalsebene für diesen Datensatz besser eignet. Jedoch kann bei der kleinen Anzahl von lediglich einer Testperson nur schwer beurteilt werden, ob sich dieser Unterschied hält oder nur ein einmaliges Phänomen vorliegt. So wurde bei den Ergebnissen der Hauptstudie deutlich, dass sich die Genauigkeit zwischen den einzelnen Testpersonen um mehr als fünf Prozentpunkte unterscheiden kann. Wird nun die Auswertung der mittleren Genauigkeit aus dem Hauptversuch herangezogen, kann festgestellt werden, dass beide Modelle mit einer sehr ähnlichen durchschnittlichen Genauigkeit von etwa 84 % bewertet werden können. Bei einer etwas genaueren Betrachtung lässt sich ein kleiner Unterschied von 0,37 Prozentpunkten angeben, der jedoch gegenüber den Schwankungen zwischen den einzelnen Testpersonen vernachlässigbar klein ist.

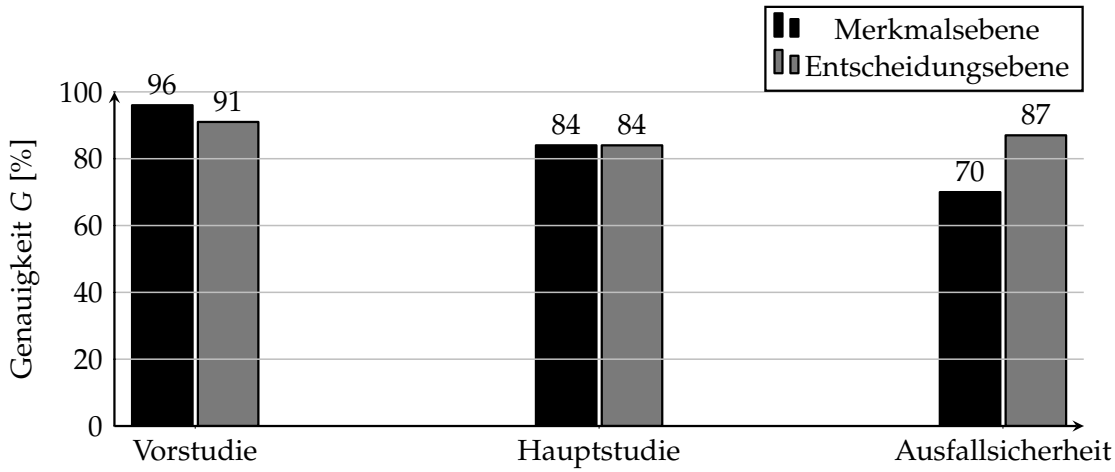


Abbildung 7.19.: Vergleich der Fusionsmethoden auf Basis durchschnittlicher Genauigkeiten

Unter Berücksichtigung der Einschätzung des Trainings der Modelle zeigt sich, dass das Modell auf Basis der Fusion auf Merkmalsebene besser für die Klassifikation des Fahrerzustands geeignet ist. Die Performanz ist recht ähnlich, wenn auch minimal kleiner als bei dem Modell auf Basis der Fusion auf Entscheidungsebene. Das Modell auf Basis der Fusion auf Merkmalsebene ist hingegen schlanker und schneller zu trainieren ist.

Zuletzt soll noch ein Blick auf den letzten Testdatensatz zur Untersuchung der Ausfallsicherheit gegeben werden. Hierbei zeigt sich der größte Unterschied dieses Vergleichs. Das Modell auf Basis der Fusion auf Entscheidungsebene schneidet hier mit 87 % in der durchschnittlichen Genauigkeit ab und ist damit um 17 Prozentpunkte besser als das Modell mit Fusion auf Basis der Merkmalsebene, welches nur 70 % erreicht. Hierbei zeigt sich der größte Vorteil der Fusion auf Entscheidungsebene, denn diese zeigt sogar ein besseres Ergebnis als bei den Testdaten des Hauptversuchs. Das Modell auf Basis der Fusion auf Merkmalsebene erreicht allerdings eine schlechtere Performanz als in dem vorher gezeigten Hauptversuch. So kann für die Auswertung über die Ausfallsicherheit gesagt werden, dass für diesen Fall das Modell mit der Fusion auf Entscheidungsebene gewählt werden sollte.

Wenn nun die Ergebnisse über alle drei Testdatensätze verglichen werden, sollte berücksichtigt werden, dass jeder Datensatz eine andere Größe aufweist. Der Testdatensatz der Vorstudie ist derjenige mit der kleinsten Größe und ist damit auch am wenigsten aussagekräftig. Der Testdatensatz der Ausfallsicherheit hat eine mittlere Größe, denn hier sind die Daten von zwei Probanden, anstatt nur von einem, wie in dem Testdatensatz der Vorstudie. Am aussagekräftigsten ist der Testdatensatz der Hauptstudie, da in diesem die Daten von drei Probanden enthalten sind. Da bei dem aussagekräftigsten Datensatz nur ein geringer Unterschied zu sehen ist, müssen die anderen Auswertungen für eine Bewertung herangezogen werden. Bei dem Vorversuch, der, wie erwähnt, weniger aussagekräftig ist, steht das Modell mit Fusion auf Merkmalsebene etwas besser da. Hierbei ist der Unterschied jedoch geringer als bei den Testdaten aus der Auswertung der Ausfallsicherheit. Bei dieser Auswertung erreicht das Modell auf Basis der Fusion auf Entscheidungsebene mit

größerem Abstand eine bessere Performanz. Da hierbei die Aussagekraft durch die Größe des Datensatzes dazu noch etwas größer ist, könnte insgesamt das Modell auf Basis der Entscheidungsebene in Bezug auf die Performanz bevorzugt werden. Gerade der Unterschied in der Auswertung zur Ausfallsicherheit zeigt den deutlichen Vorteil gegenüber dem anderen Modell.
Natürlich kommt es immer auf die Anwendung an, welches Modell am besten geeignet ist. So unterscheidet sich auch die Einschätzung in den beiden Kriterien wie Training und Performanz. Dazu könnten noch weitere Kriterien eine Rolle für eine Umsetzung spielen. So ist beispielsweise ein Kriterium, welches hier nicht untersucht wurde, die Modellgröße. Diese ist besonders relevant für eine Umsetzung auf Zielhardware. Mit der Modellgröße steigt neben dem Rechenaufwand und damit der benötigten Rechengeschwindigkeit für eine Onlineumsetzung auch die notwendige Speichergröße. Beides hat Einfluss auf die Kosten einer Zielhardware.

8

Zusammenfassung und Ausblick

In diesem Kapitel soll die vorliegende Arbeit noch einmal in ihren wesentlichen Punkten zusammengefasst werden. Zunächst wurde eine kurze Einführung in das Thema gegeben. In dieser wird der Hintergrund des automatisierten Fahrens besonders in Bezug auf Automatisierungsstufen erläutert. So geht es in der vorliegenden Arbeit im Wesentlichen um eine Automatisierung des Levels 2. Dazu erörtert die Einleitung bereits bekannte Systeme der Fahrerzustandserkennung mit Fokus auf die Müdigkeitserkennung. Hierbei sind zwei Methoden bekannt: die Erkennung über Lenkbewegungen und über eine Kamera, welche die Augen auswertet. Bei letzterer wird bereits Bezug auf den PERCLOS-Wert genommen, welcher auch als Merkmal verwendet wird. Darüber hinaus wird über Methoden berichtet, die eine Auswertung des Herzens verwenden, um die Müdigkeit zu bestimmen. Dabei werden mehrere Studien vorgestellt, die einen Zusammenhang zwischen Veränderungen in Herzdaten und Müdigkeit zeigen. Ein weiteres Thema ist die Erfassung von Müdigkeit nicht über Körperfunktionen, sondern über psychologische Versuche, die im Wesentlichen über Fragebögen die Müdigkeit aufnehmen. In der vorliegenden Arbeit wird für die Hauptstudie eine der Skalen (KSS) verwendet, welche aus der genannten Literaturanalyse hervorgeht.
Um nun tiefer in das Thema des psychologischen Fahrerzustands einsteigen zu können, wird in dem Kapitel *Psychologie und Physiologie bei längerer automatisierter Fahrt* eine Grundlage zu dieser Thematik geschaffen. Zunächst ist eine Aufteilung in verschiedene Untergruppen gegeben, wobei in der vorliegenden Arbeit besonders der kognitive Fahrerzustand von Interesse ist, der aber auch mit den anderen Untergruppen wie motorischer und sensorischer Fahrerzustand verbunden ist. Darauf wird die Situation in dem relevanten Teil des automatisierten Fahrens aus psychologischer Sicht betrachtet. Hier wird die Aufmerksamkeit als Begriff eingeordnet. Dabei ist festzuhalten, dass die Vigilanz neben der selektiven und geteilten Aufmerksamkeit mit der zentralnervösen Aktivierung verbunden ist. Dazu wird eine Definition zur Vigilanz genannt: Vigilanz ist die Fähigkeit, auf schwache und seltene Reize reagieren zu können. Ein zu beachtender Aspekt ist, dass die Zeitspanne der Aufmerksamkeitsbeobachtung bei mehr als 30 Minuten liegt. Dass der Mensch mit dieser Aufgabe Probleme besitzt, wurde während des Zweiten Weltkriegs festgestellt und die Definition wurde dann später bei arbeitswissenschaftlichen Untersuchungen verfasst. Um den Bezug zu dem automatisierten Fahren herzustellen,

werden in der vorliegenden Arbeit verschiedene Untersuchungen zur Vigilanz vorgestellt. So wurde die Vigilanz im Zusammenhang mit Automatisierung bereits im Flugwesen untersucht. Dazu ist in der Literatur eine Zusammenfassung von Studien zu finden, die verschiedene Vigilanzaufgaben in Versuchen betrachtet. Hierbei ist der größte Unterschied zu der vorliegenden Arbeit, dass die meisten genannten Veröffentlichungen die Vigilanzaufgabe unabhängig von der Fahraufgabe betrachten und nicht in diese integrieren. Eine weitere Herausforderung bei der Literaturrecherche hängt damit zusammen, dass in einigen eher technischen Beiträgen die Definition von Vigilanz sehr weit gefasst wird und unter Umständen äquivalent zur Müdigkeit betrachtet wird. In der vorliegenden Arbeit wird die Vigilanz nach der geltenden Definition aus der Psychologie betrachtet.

Bei der Auswahl der Merkmale für eine Erkennung der Vigilanz ist es wichtig Grundlagen der Psychophysiologie zu betrachten. In diesem Fall ist gemeint, wie sich psychologische Prozesse in messbaren physiologischen Signalen widerspiegeln. Dabei liegt in der vorliegenden Arbeit der Fokus auf Signale, die hohes Potenzial bieten, kontaktlos und damit unauffällig für den Fahrer gemessen zu werden. Zu einer Auflistung solcher Messmethoden wird in dem genannten Teil dieser Arbeit ein Literaturverweis gegeben. Darauf folgend stehen drei Organe im Fokus, für die unterschiedliche Untersuchungen zu Effekten in Bezug zur Psychologie näher beschrieben werden. Dabei ist besonders die Untersuchung von Merkmalen von Interesse. Bei den Organen handelt es sich um das Gehirn, das Herz und die Augen. Bei den Untersuchungen zu den am Gehirn messbaren Größen, hat sich das EEG-Verfahren durchgesetzt. Um hierbei messbare und für Maschinen verarbeitbare Daten zu erlangen, wird auf die Spannungsunterschiede am Gehirn eine Frequenzanalyse angewandt. Dabei lassen sich fünf Frequenzbänder ausmachen, die eine Aussagekraft zu psychologischen Effekten besitzen sollen. Diese Merkmale oder eine Kombination von diesen wurde bereits für Vigilanzuntersuchungen verwendet und haben eine Aussagekraft versprochen. Dem Herzen wird in der Literatur ebenfalls ein Zusammenhang zu psychologischen Zuständen zugeschrieben, welcher durch die Merkmale Herzfrequenz sowie Herzraten-Variabilität messbar gemacht werden kann. Hierbei ist die Zeitspanne für die Berechnung des Merkmals relevant, da es Untersuchungen gibt, die sehr lange Zeiträume von einem ganzen Tag oder mehreren Stunden betrachten. Für die vorliegende Arbeit wird eine Zeitspanne von 60 s verwendet, welche dahingegen recht kurz ist. Aber auch zu solch kurzen Zeitspannen gibt es Untersuchungen. Die Effekte können durch diese Merkmale in mehreren Untersuchungen zur Vigilanz messbar gemacht werden. Das letzte Organ, die Augen, bieten ebenso die Möglichkeit, Merkmale zu berechnen, welche eine psychologische Klassifikation geben können. Hierzu zählt der PERCLOS-Wert, der eine Analyse der Augenschließzeit beinhaltet. Dazu können auch Blinzelhäufigkeit oder das generelle Blickverhalten genutzt werden. Der PERCLOS-Wert wird in vielen Untersuchungen schon für Referenzmessungen der Müdigkeit verwendet, sodass ein Zusammenhang auch zur Vigilanz wahrscheinlich ist.

Nach den Erläuterungen zu den physiologischen Merkmalen, wird danach die Grundlage für die technische und automatisierte Verarbeitung dieser Merkmale beschrieben. Zunächst wird in dem Kapitel *Datenverarbeitung und Fusion mittels*

künstlichen neuronalen Netzwerken die Geschichte des maschinellen Lernens dargestellt, um die Grundelemente der Netzwerke einordnen zu können. Es folgt die Beschreibung einer speziellen Form der künstlichen neuronalen Netzwerke, den LSTM-Netzwerken. Diese bieten die Möglichkeit, Informationen über mehrere Zeitschritte zu teilen. Da in der Definition zur Vigilanz ganz klar eine zeitliche Änderung beschrieben wird, sollte auch das Fahrerzustandsmodell einen zeitlichen Zusammenhang berücksichtigen. Bei der Auswertung der Rohdaten der Kamera zur Berechnung der Merkmale wird eine spezielle Art von künstlichen neuronalen Netzwerken verwendet. Hier werden faltende neuronale Netzwerke verwendet, deren größter Vorteil gegenüber gewöhnlichen Netzwerken die zweidimensional geteilten Gewichte sind. Diese sorgen dafür, dass bei der Detektion eine örtliche Unabhängigkeit in der Bildebene entsteht. Ein Teil der künstlichen neuronalen Netzwerke in dieser Arbeit wird zur Klassifikation verwendet. So wird erläutert, was der Unterschied zwischen Klassifikation und Regression mit diesen Netzwerken ist und warum die Klassifikation gewählt wird. Der Schwerpunkt der Untersuchung in dieser Arbeit soll die Fusion von verschiedenen physiologischen Daten bilden. Dazu wird eine Einführung in die Fusionstheorie vorgestellt. Von den drei Fusionsstufen der Rohdatenebene, der Merkmalsebene und der Entscheidungsebene sind besonders die beiden zuletzt genannten Stufen von Interesse. Das Zusammenspiel und die Fusion innerhalb von künstlichen neuronalen Netzwerken wird bereits an einigen Stellen in der Literatur beschrieben. So werden Beispiele aus der Literatur gegeben, die ein ähnliches Vorgehen zur Fusion nutzen. Dabei gibt es sowohl die Fusion im Kontext der Fahrerzustandserkennung als auch im Zusammenspiel von mehreren LSTM-Netzwerken. Hierbei gibt es Hinweise zu unterschiedlichen Methoden der Fusion, welche auch in der hier vorliegenden Arbeit Anwendung finden.

Im darauf folgenden Kapitel *Machbarkeitsstudie zum Aufmerksamkeitsverhalten* wird die erste der beiden Studien vorgestellt. Diese Studie sollte dazu dienen, Erfahrungen mit psychologischen Studien zu erlangen und erste Ansätze zur Datenverarbeitung zu formen. Hierbei wird zunächst das Design dieser Studie vorgestellt. Zur Integration der Aufmerksamkeitsaufgabe in den Straßenverkehr wird in diesem Fall eine Signalentdeckungsaufgabe verwendet, die auf unterschiedlichen Schildern in einer Fahrsimulation basiert. In dieser Fahrsimulation wird eine monotone Autobahnfahrt in einem automatisierten Fahrzeug dargestellt. Die Probanden sollten so schnell wie möglich auf Schilder reagieren, die im Realverkehr nicht auf der Autobahn zu finden sind. Dazu werden vor und nach der Fahrt verschiedene Fragebögen verwendet, um auf subjektive Einschätzungen zur Müdigkeit und Anstrengung während des Versuchs zurückgreifen zu können. Bei der psychologischen Auswertung des Versuchs wird deutlich, dass die Definition zur Vigilanz nicht erfüllt werden kann. So wird der Versuch mit einer geringen Anzahl an Probanden noch einmal mit einer längeren Versuchslaufzeit durchgeführt, um eine Einschätzung zu dieser Variable zu bekommen. Dabei konnte auch bei länger Versuchszeit kein eindeutiges Ergebnis erzielt werden. Dennoch beschreiben die Probanden diese Aufgabe als ermüdend, langweilig und anstrengend.

Das darauf folgende Kapitel *Probanden-Studie zur Vigilanzuntersuchung im Fahrsimulator* beschäftigt sich mit der zweiten Studie, der vorliegenden Arbeit. Auch in diesem

Kapitel geht es zunächst um das Studiendesign, auch hier angefangen mit der psychologischen Aufmerksamkeitsaufgabe. In diesem Fall besteht die Aufgabe darin, den fahrsicheren Zustand zu überprüfen. Dabei gibt es drei verschiedene Reize, auf die der Proband reagieren soll. So begibt sich das Fahrzeug aus dem fahrsicheren Zustand, indem die Geschwindigkeitsbegrenzung missachtet wird, ein Hindernis auf der Straße vorliegt oder eine durchgezogene Linie überfahren wird. Diese Reize treten einmal am Anfang des Versuchs auf, dann ist über 20 Minuten kein Reiz zu sehen, woraufhin gegen Ende des Versuches wieder Reize auftreten. Auch in diesem Fall wird die Studie in einem Fahrsimulator durchgeführt, wobei sich die Fahrzeit über 60 Minuten erstreckt. Des Weiteren werden bei dieser Untersuchung die Probanden gebeten, Fragebögen auszufüllen. Hier wurde aus der Vorstudie gelernt, deshalb konnte bei dieser Studie ein Vergleich zwischen dem psychologischen Zustand vor der Fahrt gegenüber nach der Fahrt ausgewertet werden. Dabei wird auf bewährte Fragebögen aus der Literatur zurückgegriffen. Dazu wird eine zeitliche Selbsteinschätzung der Probanden zur eigenen Aufmerksamkeit erhoben, welche später zum Einordnen der physiologischen Daten dienen kann. Ebenso wie bei der Vorstudie, wird auch in diesem Kapitel eine psychologische Auswertung des Versuchs integriert. Die wichtigsten Ergebnisse bei dieser Darstellung sind die Auswertung zur Selbsteinschätzung der Müdigkeit, nach einem aus der Literatur bewährten Verfahren, die Auswertung der Reaktionszeiten und der ausbleibenden Reaktionen. Bei der Selbsteinschätzung der Müdigkeit durch die KSS kann ein signifikanter Anstieg von einem Skalenwert von etwa vier auf ungefähr sieben festgestellt werden. Die Auswertung der Reaktionszeiten zeigt einen signifikanten Anstieg gegen Ende des Versuchs, denn die mittlere Reaktionszeit steigt von etwa zwei auf über vier Sekunden an. Auch die ausbleibenden Reaktionen steigen im Verlauf des Versuchs an. Somit kann in diesem Fall die Definition für den Begriff Vigilanz aus den psychologischen Untersuchungen der Arbeitswissenschaft erfüllt werden. Damit ist gezeigt, dass eine solche Aufmerksamkeitsaufgabe im Fahrsimulator als Vigilanzaufgabe aufgefasst werden kann.

In dem Kapitel *Datenverarbeitung und Fusion von verschiedenen physiologischen Merkmalen* werden die Studien aus elektrotechnischer Perspektive ausgewertet. Dabei steht zunächst die Einordnung der Daten aus den beiden Studien im Vordergrund. Bei der Vorstudie werden die Daten durch eine Analyse der Gehirnaktivität eingeordnet. Bei der Hauptstudie ist es möglich, die Selbsteinschätzung der Probanden zu nutzen, um die Daten den verschiedenen Klassen zuzuordnen. Daraufhin wird die Struktur der Modelle vorgestellt, die mit der Fusion auf Merkmalsebene arbeiten. Hierbei wird zwischen der Vor- und Hauptstudie unterschieden, da die beiden Versuche unterschiedliche Datensätze erzeugt haben. So wird bei dem Hauptversuch zusätzlich zu den anderen Merkmalen auch noch der Sitzdruck aufgenommen und als Merkmal herangezogen. Ebenso unterscheidet sich die Länge der Messzeit in den beiden Studien, sodass auch verschieden lange Sequenzen für die Verarbeitung herangezogen werden können. Im Wesentlichen bestehen beide Modelle aber aus dem gleichen Grundaufbau. Die vorverarbeiteten Merkmale werden zunächst durch ein LSTM-Netzwerk analysiert, bei dem die Merkmale direkt fusioniert werden. Dieses sorgt für eine Berücksichtigung des Zeitverlaufs der Daten. Nach dieser

Verarbeitungsstufe wird ein vollständig verknüpftes Netzwerk als Klassifikator eingesetzt. Durch eine folgende softMax Schicht wird das Ergebnis des Klassifikators in eine Zeitreihe von Klassifikationsergebnissen umgerechnet. Diese Komponenten sind beim Vor- als auch Hauptversuch gleich. Bei dem Hauptversuch ist zusätzlich noch eine DropOut Schicht eingebracht, um einer Überanpassung entgegenzuwirken. Beide Netzwerke zeigen eine Art von Konvergenz beim Training, mit der gezeigt wird, dass eine Anpassung des Modells zu einer Verbesserung des Trainingsfehlers führt. Nach der Beschreibung der Modellstruktur auf Basis der Fusion auf Merkmalsebene, wird die Struktur der Modelle auf Basis der Fusion auf Entscheidungsebene vorgestellt. Hier ist der Ansatz, dass für jeden der Sensoren ein unabhängiges Klassifikationsnetzwerk herangezogen wird. Die Ergebnisse der einzelnen Teilmodelle werden dann durch eine Fusion zusammengeführt, sodass auch hier eine Klasse pro Zeitschritt berechnet werden kann. Für die Teilmodelle der verschiedenen Sensoren wird wieder die Struktur der Modelle verwendet, die für das Verfahren auf Basis der Fusion auf Merkmalsebene herangezogen wurde. Auch in diesem Fall ist bei dem Trainingsfehler erst eine Abnahme und dann eine Konvergenz zu sehen, sodass auch hier das Modell auf die Daten angepasst sein sollte.

Im Kapitel *Modell zur Klassifikation von Fahreraufmerksamkeit mit Fokus auf Vigilanz* wird als nächstes die Performanz der unterschiedlichen Modelle auf noch nicht ausgewerteten Testdaten bestimmt und später verglichen. Hierzu werden zunächst die Auswertungskriterien genauer erklärt. In der vorliegenden Arbeit werden für die Bewertung vier Bewertungskriterien herangezogen: Genauigkeit G, Verzugszeit t_v, Präzision P und Ausfallquote R. Anschließend werden die ersten Ergebnisse des Modells auf Basis der Fusion auf Merkmalsebene für den Testdatensatz aus der Vorstudie gezeigt. Nachdem eine hohe Genauigkeit bei den Trainingsdaten festgestellt werden konnte, wird auch bei den Testdaten eine Genauigkeit von 96 % erreicht. Ebenso wird das äquivalente Modell des Hauptversuchs mit dem Testdatensatz ausgewertet. Dieser besteht nicht nur aus den Daten eines Probanden, sondern von drei verschiedenen Probanden. Hier wird deutlich, dass die Performanz nicht bei allen Probanden aus dem Testdatensatz gleich gut ausfällt. Die Daten von einem Probanden zeigen deutlich schlechtere Performanzwerte als die anderen beiden. Bei der Auswertung des Modells auf Basis der Fusion auf Entscheidungsebene wird zunächst die Performanz der einzelnen Teilmodelle ausgewertet. Bei dem Modell der Vorstudie gibt es zwei Teilmodelle, jeweils eines für die Daten des Herzsensors und eines für die Daten der Kamera. Beide Teilmodelle zeigen eine ähnliche Genauigkeit von über 80 %. Daraufhin wird die Performanz des gesamten Modells ausgewertet. Hierbei wird eine Genauigkeit von 91 % mit dem Testdatensatz aus der Vorstudie erreicht. Danach werden die Ergebnisse aus der Hauptstudie ausgewertet. Hier wird auch zunächst die Performanz der einzelnen Teilmodelle dargestellt. In diesem Fall sind drei Teilmodelle auszuwerten, denn hier kommt das Modell der Sitzdruckdaten hinzu. Zwischen den Teilmodellen der Kameradaten und Herzdaten ist ein Unterschied festzustellen, da das Teilmodell der Herzdaten auf den Testdaten schlechtere Performanz zeigt als das Teilmodell der Kameradaten. Das Gesamtmodell, bei dem alle drei Teilmodelle kombiniert werden, erreicht bei dem ersten Testdatensatz T_1 eine Genauigkeit von etwa 75 %, bei dem Testdatensatz T_2 eine Genauigkeit von 94 % und bei dem

dritten Testdatensatz T_3 eine Genauigkeit von 84 %. Bei allen drei Werten gibt es die Erkenntnis, dass die Genauigkeit in jedem der drei Fälle über der durchschnittlichen Genauigkeit der drei Teilmodelle liegt. Bei zwei der drei Testdatensätze ist der Unterschied mit 15 Prozentpunkten recht deutlich ausgeprägt. Des Weiteren wird in der vorliegenden Arbeit noch die Ausfallsicherheit des Netzwerks analysiert. Hierzu werden zwei Datensätze herangezogen, welche zuvor beim Training aussortiert wurden, da einzelne Sensoren für eine gewisse Zeit ausgefallen sind. Weil aber genau dies in der Realität möglich ist, werden ebenfalls diese Datensätze für die Auswertung herangezogen. Bei beiden Datensätzen wird deutlich, dass das Modell mit der Fusion auf Merkmalsebene keine wirklich sinnvolle Klassifikation leisten kann. Einer der beiden Zustände wird fast über den kompletten Zeitverlauf beibehalten. Im Gegensatz dazu verändert sich die Performanz des Modells auf Basis der Fusion auf Entscheidungsebene nur geringfügig. Bei dem ersten Testdatensatz F_1 kann hiermit eine Genauigkeit von etwa 93 % und bei dem zweiten Testdatensatz F_2 eine Genauigkeit von etwa 82 % erreicht werden. Zuletzt wird in diesem Kapitel im zweiten Teilkapitel eine Bewertung zwischen den Modellen gegeben. Hierbei werden die Modelle in Bezug zum Training und zur Performanz verglichen. Der wichtigste Punkt beim Training ist die Menge der zu trainierenden Variablen. Diese ist bei dem Modell auf Basis der Fusion auf Entscheidungsebene leicht höher. Bei der Performanz werden die durchschnittlichen Genauigkeiten über alle Testdatensätze bestimmt und dann die beiden Modelle gegenübergestellt. Bei dieser Darstellung wird deutlich, dass für die Testdaten aus dem Vor- und Hauptversuch kaum Unterschiede bestehen. Insgesamt erreicht das Modell auf Basis der Merkmalsfusion auf diesen Datensätzen sechs Prozentpunkte mehr Genauigkeit. Dahingegen wird bei der Untersuchung der Ausfallsicherheit ein größerer Unterschied deutlich. Die Genauigkeit ist bei dem Modell auf Basis der Fusion auf Entscheidungsebene um 17 Prozentpunkte besser. Somit wird festgehalten, dass im Hinblick auf die Performanz das Modell auf Basis der Fusion auf Entscheidungsebene empfohlen wird. Steht eine schnelle Trainingsphase im Fokus, sollte das Modell auf Basis der Fusion auf Merkmalsebene besser geeignet sein.

Ausblick

Darüber hinaus bleiben noch einige Herausforderungen, die im Rahmen der vorliegenden Arbeit nicht bewältigt werden können. Ein wichtiger Teil ist dabei die psychologische Komponente dieser Arbeit. Im Falle der Vorstudie wurden die Daten anhand der Gehirnaktivität eingeordnet. Dabei gibt es keine weiteren Untersuchungen, inwieweit die dort gezeigte Änderung der Gehirnaktivität mit dem psychologischen Zustand zusammen hängt. Hierzu könnten psychologische Studien erfolgen, in denen die Probanden mit einer aus der Literatur bekannten Vigilanzaufgabe konfrontiert werden und dabei dann die Gehirnaktivität aufgezeichnet wird. So könnte der Zusammenhang zwischen der Vigilanz und der Gehirnaktivität noch verdeutlicht werden. Zudem könnten Experten zusätzlich den Zeitverlauf der Gehirnaktivität bewerten, wie es bei klinischen Studien üblich ist. Ein weiterer Schritt könnte dann mit einer Vigilanzaufgabe im Kontext des automatisierten Fahrens erfolgen.

Auch die im Hauptversuch verwendete Methode zur Einordnung der Daten wird in der vorliegenden Arbeit nicht verifiziert. Hier sollte eine psychologische Studie erfolgen, die eine Auswertung über die Selbsteinschätzung untersucht. Hier könnte eine Kombination zu der vorher genannten Studie sinnvoll sein, da ein objektives Maß für die Selbsteinschätzung notwendig ist. Dies könnte durch die Analyse der Gehirnaktivität erfolgen, welche auch die zeitliche Selbsteinschätzung bewerten könnte. Auch hier könnte ein Rückschritt auf eine aus der Literatur bekannten Aufgabe sinnvoll sein, um die unterschiedlichen Fragen zur zeitlichen Selbsteinschätzung genauer beleuchten zu können. Später sollten auch Versuche in dem Umfeld des automatisierten Fahrens erfolgen, um eine Übertragbarkeit auf den in der vorliegenden Arbeit vorgestellten Versuch sicherzustellen. Besonders eine Information darüber, wie zeitlich genau Probanden sich selber einschätzen können, ist von Interesse, da diese Information bei einer folgenden Bewertung von Daten wichtig ist. So könnten die psychologische Situation für gewisse Zeitfenster dann gleich bewertet und die zeitliche Auflösung herabgestuft werden.
Wie die Untersuchung in der Vorstudie der vorliegenden Arbeit gezeigt hat, ist es nicht einfach, abzuschätzen, welche Signale geeignet für eine Vigilanzaufgabe sind. Die Definition aus den psychologischen Versuchen beschreibt schwache und seltene Reize. Zu der Häufigkeit der Reize gibt es bereits Metastudien, die diese Zahl aus verschiedenen Vigilanzuntersuchungen vergleichen und somit ein übliches Maß angeben. Zu der Signalstärke existiert keine einfache Vergleichsmöglichkeit, sodass diese nicht aus der Literatur zu bestimmen ist. Hier könnten in Zukunft noch weitere psychologische Studien folgen, die eine Signalstärke im Straßenverkehr untersuchen. Dabei könnte zum Beispiel im Fokus stehen, welche Art von Hindernissen einen starken oder schwachen Reiz beim Fahrer auslösen. Auch die anderen Reize, die in der Hauptstudie der vorliegenden Arbeit herangezogen wurden, sollten noch tiefgreifender untersucht werden.
Darüber hinaus sollte der Datensatz aus der vorliegenden Arbeit noch durch mehr Untersuchungen gleicher Art erweitert werden. Die Anzahl der Probanden ist in diesem Fall recht gering, sodass noch keine gute Abbildung der Allgemeinheit in dem Datensatz möglich ist. So sollte diese Studie auch besonders auf weibliche und diverse Teilnehmer ausgeweitet werden. Ebenso sind keine großen Unterschiede in der Ethnie gegeben, sodass die Studie insgesamt diverser werden sollte. So könnten physiologische Unterschiede einen großen Einfluss auf den Ansatz, der hier vorliegenden Arbeit hervorbringen. Besonders der verwendete Kamerasensor ist hiervon betroffen, da Farbbilder analysiert werden. Durch eine insgesamt größere Datenmenge könnte dazu der Gradient beim Training der Modelle besser geglättet werden, sodass sich zudem die Performanz erhöhen könnte. Auch die statistische Auswertung der psychologischen Ergebnisse wird durch eine größere Stichprobe genauer und zuverlässiger.
In der vorliegenden Arbeit haben alle Versuche in einem statischen Fahrsimulator stattgefunden. Auch in diesem Aspekt könnte der Realismus der Studie erhöht werden. So sollten folgende Studien in einem dynamischen Fahrsimulator stattfinden. Die zusätzliche Rückmeldung der Fahrdynamik an den Fahrer könnte einen weiteren Reiz ausmachen, der in der vorliegenden Arbeit nicht aufgetreten ist. Besonders in

Bezug auf die Geschwindigkeitsänderung, die hauptsächlich für die Auswertung der Vigilanz verwendet wurde, könnte hierdurch vom Probanden verändert wahrgenommen werden. Ein weiterer Schritt, um besser automatisierte Fahrten darzustellen, könnte durch ein Wizard-of-Oz-Fahrzeug erfolgen. Dabei findet der Versuch in einem Realfahrzeug statt, welches von einem Versuchsleiter gesteuert wird, wobei dem Probanden vermittelt wird, dass es automatisiert fährt. So könnte unter kontrollierten Bedingungen eine Vigilanzuntersuchung auf der Straße stattfinden. Neben den zusätzlichen Reizen, die der Proband erfahren könnte, kommt hinzu, dass der Proband auch weitere Emotionen entwickeln könnte. So könnte zum Beispiel in einem Realfahrzeug durch Angst die Herzfrequenz deutlich verändert sein. Dies könnte Einfluss auf die Auswertung der Herzdaten haben. Auch die Beobachtung der Augen könnte unter real wechselnden Einflüssen erschwert sein. Das Vertrauen in die Automatisierung spielt ebenfalls eine Rolle bei den Emotionen, die Einfluss auf die Fahreraufmerksamkeit haben. Diese Abhängigkeit könnte durch einen Fragebogen vor der Fahrt bereinigt werden. So könnte die Gruppe der Probanden in mehrere Untergruppen unterteilt werden, die als Kriterium das Vertrauen zu der Automatisierung berücksichtigt.

Neben der Vigilanz könnten mit dem in der hier vorliegenden Arbeit beschriebenen Verfahren auch noch andere Zustände des Fahrers ermittelt werden. Gerade im Kontext von höheren Automatisierungsstufen (SAE Level 3) könnte für das Fahrzeug interessant werden, ob der Fahrer schläft oder nur die Augen geschlossen hat. Hierzu sollten mit dem Sensorsystem Versuche durchgeführt werden, bei denen die Probanden schlafen können. Durch die Messung der Gehirnaktivität sollte festgestellt werden können, in welchem Zustand sich der Fahrer befindet. Erste Untersuchungen in eine solche Richtung haben bereits mit einzelnen Sensoren aus dem Sensorsetup der vorliegenden Arbeit stattgefunden. So wurden mit dem Gehirnaktivitätsmessgerät Daten von Probanden aufgenommen, die eine gewisse Zeit ihre Augen geöffnet haben sollten und danach eine gewisse Zeit die Augen geschlossen halten sollten. Diese Daten wurden nicht tiefgreifend untersucht, jedoch konnte zwischen diesen beiden Zuständen bereits ein Effekt in der Gehirnaktivität festgestellt werden. Darüber hinaus wurde versucht, mit dem kompletten Sensorsystem eine schlafende Person aufzunehmen. Jedoch ist es schwierig, Probanden unter normalen Bedingungen in einer solchen Situation tatsächlich zum Schlafen zu bringen. Dies ist im Verlauf der vorliegenden Arbeit nicht gelungen.

Durch die zuvor beschriebene kleine Untersuchung der Ausfallsicherheit wird deutlich, dass es zielführend wäre, das Sensorsystem robuster zu gestalten. Dies könnte dadurch gegeben werden, dass das gleiche Merkmal von mehreren Sensoren erfasst wird. Gerade bei der Aufnahme des Herzens sind hier schon neue Verfahren in der Literatur beschrieben. Eine weitere Möglichkeit, die Herzdaten zu erfassen, könnte durch einen Radarsensor bestehen. Wie auch von [Leo+18] beschrieben, ist es möglich, durch einen solchen Sensor den Abstand zur Herzwand zu bestimmen, sodass jeder Herzschlag des Fahrers erfasst werden könnte. So zeigt beispielsweise auch [Sch+18] einen Versuchsaufbau, um eine solche Messung zu realisieren. Eine andere Möglichkeit, eine Messung des Herzens vorzunehmen, könnte durch ein optisches Verfahren stattfinden. Für Farbbilder von normalen Kameras wurde schon

von [Wu+12] ein Verfahren vorgestellt, bei dem einzelne Herzschläge deutlich sichtbar gemacht werden. In Kombination mit der Analyse des Gesichts, die in der vorliegenden Arbeit bereits praktiziert wird, könnten mit dieser Methode die Herzdaten aus den Farbbildern extrahiert werden. Fraglich ist hierbei noch, wie stabil ein solches Verfahren gegenüber Umwelteinflüssen und Bewegungsartefakten ist. Hierzu sollten also in Zukunft noch Untersuchungen stattfinden, welche eine Adaption des vorgestellten Verfahrens in das Fahrererkennungsproblem erfassen. Gerade die Beobachtung des Fahrers auch bei Nacht könnte hierbei große Unterschiede ergeben. Dazu könnte untersucht werden, ob eine aktive Beleuchtung des Fahrers notwendig wäre und ob diese als störend empfunden wird. Dazu könnte noch herausgefunden werden, in welcher Weise eine Beleuchtung mit Licht im nicht sichtbaren Wellenbereich eine ähnliche Auswertung ermöglicht.

Ein weiterer Punkt, der in zukünftigen Untersuchungen erforscht werden sollte, ist die Ausführung der Fahrerzustandserkennung in Echtzeit. Alle in der vorliegenden Arbeit gezeigten Ergebnisse sind offline entstanden. Ein wesentlicher Bestandteil der Ausführungszeit entsteht schon bei der Auswertung der Kamerabilder. In dieser Arbeit fiel die Entscheidung auf eine Auswertung, bei der die Genauigkeit mehr im Fokus stand als die Ausführungszeit. So lief die Berechnung auf einer recht potenten Hardware nur mit einer Geschwindigkeit von fünf Bildern pro Sekunde. Auch die komplette Signalverarbeitungskette wurde nicht direkt hintereinander ausgeführt. Denkbar wäre auch eine Untersuchung, ob die Auswertung auch mit kürzeren Zeitreihen funktioniert. Mit dem Datensatz aus der vorliegenden Arbeit wurde ein solches Vorgehen kurz getestet. Hierbei wurde der Fahrerzustand mit einer Zeitreihe von zehn Minuten berechnet und dann mit einer Art Fensterfunktion über die komplette Versuchszeit verschoben. Dieses Vorgehen zeigte jedoch eine schlechtere Genauigkeit und wurde nicht im Detail ausgewertet. Ein solches Vorgehen würde die Anzahl der Berechnungen in dem Modell und damit auch dessen Laufzeit reduzieren. Generell müsste bei allen Stufen der Datenverarbeitung ein Kompromiss zwischen Genauigkeit und Laufzeit untersucht werden. Zudem könnte das Training der Modelle noch verbessert werden, denn in der vorliegenden Arbeit wurden die Trainingsparameter vergleichbar gehalten. Eine Optimierung der verschiedenen Hyperparameter könnte zu einer Verbesserung der Trainingsphase führen. Dadurch könnte das Training schneller werden und die Genauigkeit der Modelle noch weiter erhöht werden.

Wenn eine Berechnung in Echtzeit erfolgreich implementiert und erprobt werden kann, könnte das berechnete Fahrerzustandssignal weiter genutzt werden. Hierbei ist ein Gedanke, ein übergeordnetes Fahrermodell zu erstellen. So gibt es neben der Vigilanz beziehungsweise dem kognitiven Zustand auch noch die weiteren Fahrerzustände, den motorischen und sensorischen Zustand. Alle Zustände könnten in diesem überlagerten Modell dargestellt und deren Interaktion untereinander abgebildet werden. Ein anderes Beispiel ist aus den Fahrsimulationen bekannt. Hier gibt es sogenannte Agenten, die ein Fahrverhalten von anderen Verkehrsteilnehmern darstellen sollen. Diese Agenten sind mit gewissen Parametern parametrisiert, wie zum Beispiel einem Wert für die Reaktionszeit. Diese bereits bekannten Modelle könnten auch für ein Fahrermodell herangezogen werden, sodass das Auto den

Zustand des Fahrers in die Zukunft simulieren kann. Bestimmte Parameter dieser Modelle könnten durch das in der vorliegenden Arbeit entwickelte Verfahren in gewissem Maße bestimmt werden. Somit könnten die Parameter in dem überlagerten Modell aktualisiert werden. So könnten die Modelle für die Prädiktion des Fahrerzustands genauer gemacht werden.
Für den Fall, dass eine echtzeitfähige Fahrerzustandserkennung mit oder ohne Integration in ein Fahrermodell erstellt wurde, kann das resultierende Signal weiter genutzt werden. Hierbei kann eine Idee sein, auf einen unerwünschten Fahrerzustand zu reagieren, indem eine Aktion versucht, den Fahrer zu beeinflussen. Die einfachste Umsetzung einer solchen Aktion wäre die Rückmeldung des gemessenen Zustands als visuelle Meldung direkt an den Fahrer. In der Literatur würde ein solches Vorgehen als Neurofeedback bezeichnet werden, da der psychologische Zustand für den Fahrer sichtbar gemacht wird. Dies wird auch in den Serienfahrzeugen bei der Müdigkeitserkennung angewendet. Bei dem psychologischen Zustand der vorliegenden Arbeit ist ebenfalls eine solche Anzeige denkbar. Jedoch steht noch eine Untersuchung aus, wie der Fahrer mit der Information, dass er unaufmerksam ist, umgeht. Neben der visuellen Anzeige sind auch noch andere Aktionen des Systems denkbar. Beispielsweise könnte in Kombination zu dem visuellen Hinweis oder auch alleinstehend ein Warnton ausgegeben werden. Ebenso ist ein haptisches Feedback denkbar, wie ein Vibrieren des Sitzes oder das Anziehen des Gurtes. Die Fragestellung sollte sein, welches dieser Signale am besten auf den psychologischen Zustand Vigilanz einwirken kann. Hier könnte in einem Versuch zwischen den verschiedenen Möglichkeiten und einer Fahrt ganz ohne Feedback die Aufmerksamkeit verglichen werden. Neben der Aktivierung des Fahrers durch ein Signal könnte auch der Ursache der Vigilanz gezielt entgegengewirkt werden. Aus der Psychologie ist bekannt, dass gerade die Vigilanz eher bei einer mentalen Unterforderung auftritt, sodass diese durch eine Aktion erhöht werden könnte. Da das in der vorliegenden Arbeit untersuchte Szenario auf einer automatisierten Fahrt basiert, könnte eine Aktion die Änderung der Automatisierung sein. Hierbei könnten dem Fahrer wieder mehr Aufgaben zugeteilt werden, sodass dieser nicht mehr im gleichen Maße unterfordert ist. Eine andere Art, die Aktivierung des Fahrers hochzuhalten, wird von [Mar+18] beschrieben. In diesem Fall liegt der Fokus auf einer automatisierten Fahrt des SAE Level 3, bei dem die Aufmerksamkeit nicht mehr auf dem Fahren liegen muss und so die Aufgaben auch nicht auf die Fahrzeugführung bezogen werden müssen. Anders als in der genannten Untersuchung, sollte die Aufgabe auf die Fahrt bezogen sein. Vorstellbar wäre für diesen Fall ein bestimmtes Blickmuster abzufordern wie den Blick in verschiedene Spiegel. Auch eine Kombination mit Lichtpunkten oder Head-Up-Displays, welche die Aufmerksamkeit des Fahrers leiten könnten, wären eine Untersuchung wert. In jedem Fall muss untersucht werden, wie sich die Aktion auf die Vigilanz des Fahrers auswirkt und ob es überhaupt möglich ist, in hohem Maße Einfluss auf die Vigilanz zu nehmen.
Letztendlich bleibt in dem Feld der Fahrerbeobachtung im Fokus auf Vigilanz noch Potential für nachfolgende Untersuchungen. Mit dem in der vorliegenden Arbeit vorgestellten Modellen ist ein erster Schritt für diese Forschung absolviert.

A

Anhang

A.1. Messsystem zur Aufnahme von physiologischen Daten

A.1.1. Sensoraufbau

Um die in Abschnitt 2 beschriebenen Körperfunktionen erfassen zu können, wird ein Sensorsystem benötigt. Ein Überblick über mögliche Sensoren in einem Fahrzeug wird von Leonhardt, Leicht und Teichmann [Leo+18] gegeben, wobei in dem vorliegenden Beitrag der Fokus auf kontaktlose Sensoren gelegt wird. Bei Sahayadhas, Sundaraj und Murugappan [Sah+12b] werden verschiedene Messmethoden für die Müdigkeit zusammengetragen, sodass sich auch in dem vorliegenden Beitrag der Sensoraufbau an diesen Beispielen orientiert.
Bei der hier vorgestellten Studie wird eine Kombination aus verschiedenen Sensoren verwendet. Dabei stehen sowohl kontaktlose als auch kontaktbehaftete Sensoren zur Messung der physiologischen Merkmale aus Abschnitt 2 zur Verfügung. Es werden zwei Kameras auf dem Armaturenbrett, ein Sensor am Brustkorb und ein Sensor an der Stirn des Probanden in dem Sensoraufbau verwendet.

Lenkrad

Zur Messung der Reaktionszeiten wird der Knopfdruck an dem Lenkrad verwendet. Um zu bestimmen, worauf der Proband reagiert, wird die Startzeit des Videos verwendet, sodass ein einzelnes Bild des Videos mit dem Zeitpunkt des Knopfdrucks in Zusammenhang gebracht werden kann. Um nun die Reaktionszeiten zu bestimmen, wird das Video in Einzelbilder aufgeteilt. Daraufhin wird das Bild aus den Einzelbildern bestimmt, auf dem das Straßenschild klar von anderen Straßenschildern unterschieden werden kann, welches dann als Referenzbild und Zeitpunkt für die Zeitmessung dient. Zuletzt wird die Differenz zwischen dem Zeitpunkt des Knopfdrucks und dem Zeitpunkt vom Erscheinen des Referenzbilds gebildet.

Elektroenzephalogramm

Für die Messung der Gehirnaktivität wird ein Elektroenzephalogramm (EEG) benutzt. Von Hecht, Feldhütter, Radlmayr, Nakano, Miki, Henle und Bengler

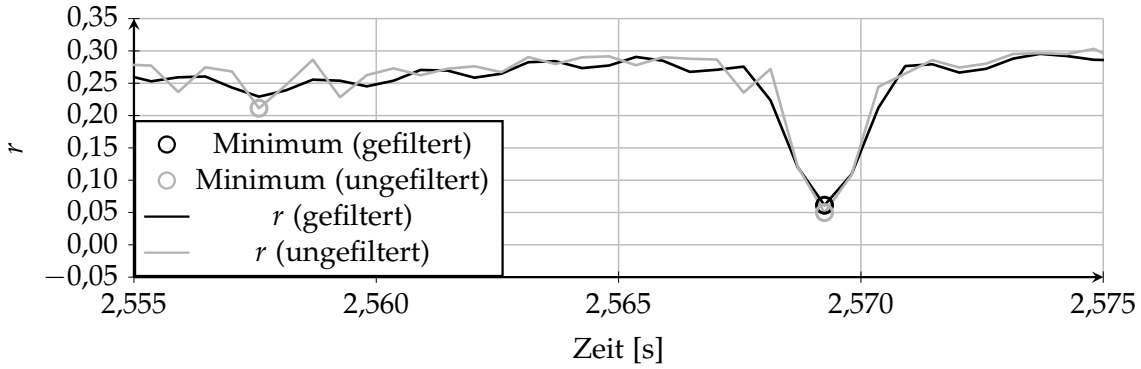


Abbildung A.1.: Bestimmung der Augenschließung bei gefiltertem und ungefiltertem Signal

[Hec+19] wird beschrieben, dass diese Messung die besten Ergebnisse für eine Fahrerzustandsklassifikation gibt. Dabei wird darauf hingewiesen, dass der Sensor kontaktbehaftet ist und zusätzlich oftmals ein Kontaktgel verwendet wird, sodass der Proband abgelenkt werden könnte. Um bei der vorliegenden Studie die Messung so nichtinvasiv wie möglich zu halten, wird ein kabelloses Stirnband mit trockenen Elektroden benutzt. Das verwendete Stirnband ist mit vier Elektroden und einer Referenzelektrode ausgestattet. Dies ist ausreichend, da keine qualitative Aussage über die örtliche Herkunft der Gehirnaktivität gegeben werden soll. Wenn eine starke Änderung der Gehirnaktivität vorliegt, dann kann von verschiedenen Zuständen des Gehirns ausgegangen werden, wie X. Zhang u. a. [Zha+17b] schon für wenige Elektroden zur Müdigkeitserkennung von Zugfahrern zeigt und verifiziert. Zudem bietet der verwendete Sensor noch weitere Messwerte wie Gyroskop-, Beschleunigungswerte und eine Photoplethysmografie (Herzschlag).

Elektrokardiogramm

Da ein Elektrokardiogramm über elektrische Signale des Herzens gemessen wird, muss dieses mit einem kontaktbehafteten Sensor durchgeführt werden. Hier wird ebenso wie bei dem EEG ein kabelloser Sensor verwendet, um keine Störungen durch Verkabelungen zu erhalten. Dieser Sensor misst sowohl die Herzschläge pro Minute, als auch die Zeiten zwischen den einzelnen Herzschlägen (R-R-Intervallen). Die Messwerte werden mit 1 Hz ausgegeben, wobei die Zeit zwischen einzelnen Herzschlägen gegebenenfalls unter eine Sekunde betragen kann, sodass mehrere Messwerte gleichzeitig ausgegeben werden können.

Kamera

Um eine kontaktlose Beobachtung des Fahrers zu erhalten, werden zwei Kameras verwendet. Eine ist hinter dem Lenkrad und die zweite an der A-Säule auf der Beifahrerseite auf dem Armaturenbrett befestigt. Die Kameras nehmen Farbbilder mit einer Auflösung von 1920 mal 1080 Pixel mit 30 Bildern pro Sekunde auf. Um die Informationen über das Auge zu extrahieren, wird im ersten Schritt der von Z. Cao, G. Hidalgo Martinez, T. Simon, S. Wei und Y. A. Sheikh [Z C+19] beschriebene

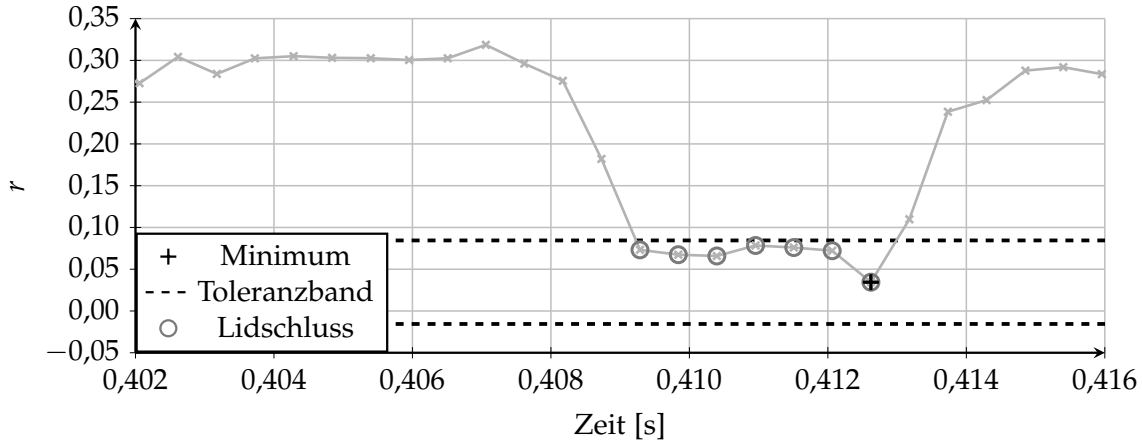


Abbildung A.2.: Höhen- zu Breitenverhältnis für eine Augenschließung und Bestimmung der Schließzeit

Algorithmus verwendet. Dieser liefert Schlüsselpunkte der Augen in Pixelkoordinaten, mit denen dann die Breite des linken Auges b_l und des rechten Auges b_r, sowie die Höhe des linken Auges h_l und des rechten Auges h_r, jeweils als Pixelabstand, bestimmt werden können. Aus diesen Werten wird dann wie folgt das Höhen- zu Breitenverhältnis r eines Datenpunkts bestimmt:

$$r = \frac{\frac{h_l}{b_l} + \frac{h_r}{b_r}}{2} \tag{A.1.1}$$

Beispielhaft ist in Abbildung A.1 das Höhen- zu Breitenverhältnis r eines Probanden abgebildet, wobei in diesem Fall der Wert bei geöffnetem Auge etwa 0,3 beträgt und bei geschlossenem Auge bei etwa 0,07 liegt. Um Messausreißer aus dem Signal r zu entfernen, werden alle Messwerte, die mehr als 0,05 nach oben, da nur Augenschließungen in diesem Fall interessant sind, von dem Mittelwert (in dem gezeigten Fall $\varnothing r = 0,27$) abweichen auf den Mittelwert gesetzt. Dieses Signal ist jedoch immer noch verrauscht, sodass ein Tiefpassfilter mit einer Grenzfrequenz von 15 Hz (halbe Abtastfrequenz) angewendet wird, welches durch die graue Linie in der Abbildung A.1 visualisiert wird. Dieses Filter dient dazu besser zeitliche Minima zu finden, um eine Augenschließung von Rauschen zu unterscheiden, wie bei 2,557 s sichtbar wird. Denn hier sorgt das Rauschen für eine Fehlerkennung, welche bei dem gefilterten Signal nicht auftritt. Ebenso wäre ein gleitender Mittelwert zur Glättung des Signals geeignet, wobei auch hier ein Parameter mit der Fensterlänge des Filters einzustellen wäre. Dazu werden nur Minima mit einer Schartenhöhe von mindestens 0,07 als Augenschließung gezählt, sodass keine Nebenminima durch das Rauschen hinzukommen.
Um die Schließdauer zu bestimmen, werden von dem Minimum nach links und rechts alle Messpunkte markiert, die in einem Toleranzband von 0,05, wie in der Abbildung A.2 gezeigt, um den Wert des Minimums liegen. Zuletzt wird noch die Anzahl dieser markierten Messpunkte durch die Anzahl der Messungen in dem Intervall von 60 s (wie in Abschnitt 2 beschrieben) bestimmt, um die prozentuale Schließdauer, also den PERCLOS-Wert, zu erhalten.

Sitzdruck



Abbildung A.3.: Anordnung der Sitzdrucksensoren mit den Sitzdrucksensoren SP1 bis SP7 in der Sitzfläche und SP8 bis SP12 in der Rückenlehne

A.2. Vigilanzstudie

Fragebogen vor der Fahrt:

Information Classification: Internal

Teil 1:
Bitte beschreiben Sie wie jedes Wort ***gerade*** ihren Gefühlszustand widerspiegelt:

	gar nicht	ein wenig	etwas	sehr	äußerst
1. Unzufrieden	o	o	o	o	o
2. Aufmerksam	o	o	o	o	o
3. Niedergeschlagen	o	o	o	o	o
4. Traurig	o	o	o	o	o
5. Aktiv	o	o	o	o	o
6. Ungeduldig	o	o	o	o	o
7. Genervt	o	o	o	o	o
8. Sauer	o	o	o	o	o
9. Irritiert	o	o	o	o	o
10. Grantig	o	o	o	o	o

Teil 2:
11. Ich bin gerade entschlossen meine Leistungsziele zu erreichen:

o	o	o	o	o
gar nicht	ein wenig	etwas	sehr	äußerst

12. Ich will die folgende Aufgabe erfolgreich bewältigen:

o	o	o	o	o
gar nicht	ein wenig	etwas	sehr	äußerst

13. Ich bin gerade motiviert die Aufgabe zu erledigen:

o	o	o	o	o
gar nicht	ein wenig	etwas	sehr	äußerst

14. Ich versuche im Moment mich selbst kennenzulernen:

o	o	o	o	o
gar nicht	ein wenig	etwas	sehr	äußerst

15. Ich denke gerade viel über mich selbst nach:

o	O	o	o	o
gar nicht	ein wenig	etwas	sehr	äußerst

16. Ich neige momentan zu Tagträumen:

o	o	o	o	o
gar nicht	ein wenig	etwas	sehr	äußerst

17. Ich bin mir aktuell über meine Fähigkeiten bewusst:

o	o	o	o	o
gar nicht	ein wenig	etwas	sehr	äußerst

18. Ich fühle mich im Moment Selbstbewusst:

o	o	o	o	o
gar nicht	ein wenig	etwas	sehr	äußerst

19. Ich mache mir Gedanken darüber, was andere gerade über mich denken:

o	o	o	o	o
gar nicht	ein wenig	etwas	sehr	äußerst

20. Ich bin aktuell besorgt über den Eindruck, den ich mache:

o	o	o	o	o
gar nicht	ein wenig	etwas	sehr	äußerst

21. Ich erwarte, dass ich bei dieser Aufgabe gut abschneide:

o	o	o	o	o
gar nicht	ein wenig	etwas	sehr	äußerst

22. Im Moment habe ich das Gefühl, alles unter Kontrolle zu haben:

o	o	o	o	o
gar nicht	ein wenig	etwas	sehr	äußerst

23. Ich habe darüber nachgedacht, wie andere bei dieser Aufgabe abgeschnitten haben:

o	o	o	o	o
gar nicht	ein wenig	etwas	sehr	äußerst

24. Ich habe darüber nachgedacht, wie ich mich fühlen würde, wenn mir gesagt werden würde, wie ich bei dieser Aufgabe abschneide:

o	o	o	o	o
gar nicht	ein wenig	etwas	sehr	äußerst

Karolinska Sleepiness Scale (KSS) vor der Fahrt:

KSS1. Wählen Sie bitte diejenige Aussage, die am besten den Grad ihrer Schläfrigkeit (bzw. Wachheit) beschreibt:

- o Extrem wach
- o Sehr wach
- o Wach
- o Ziemlich wach
- o Weder wach noch schläfrig
- o Einige Anzeichen von Schläfrigkeit
- o Schläfrig, aber kann noch ohne Mühe wach bleiben
- o Schläfrig, habe Mühe wach zu bleiben
- o Sehr schläfrig, kann nur mit großer Mühe wach bleiben; kämpfe gegen den Schlaf

KKS und Fragebogen nach der Fahrt:

KSS2. Wählen Sie bitte diejenige Aussage, die am besten den Grad ihrer Schläfrigkeit (bzw. Wachheit) beschreibt:

- o Extrem wach
- o Sehr wach
- o Wach
- o Ziemlich wach
- o Weder wach noch schläfrig
- o Einige Anzeichen von Schläfrigkeit
- o Schläfrig, aber kann noch ohne Mühe wach bleiben
- o Schläfrig, habe Mühe wach zu bleiben
- o Sehr schläfrig, kann nur mit großer Mühe wach bleiben; kämpfe gegen den Schlaf

Teil 1:

Bitte beschreiben Sie wie jedes Wort ihren Gefühlszustand ***während der Aufgabe*** widerspiegelt:

	gar nicht	ein wenig	etwas	sehr	äußerst
1. Unzufrieden	o	o	o	o	o
2. Aufmerksam	o	o	o	o	o
3. Niedergeschlagen	o	o	o	o	o
4. Traurig	o	o	o	o	o
5. Aktiv	o	o	o	o	o
6. Ungeduldig	o	o	o	o	o
7. Genervt	o	o	o	o	o
8. Sauer	o	o	o	o	o
9. Irritiert	o	o	o	o	o
10. Grantig	o	o	o	o	o

Teil 2:

11. Ich war entschlossen meine Leistungsziele zu erreichen:

o	o	o	o	o
gar nicht	ein wenig	etwas	sehr	äußerst

12. Ich wollte bei der Aufgabe erfolgreich sein:

o	o	o	o	o
gar nicht	ein wenig	etwas	sehr	äußerst

13. Ich war motiviert die Aufgabe zu erledigen:

o	o	o	o	o
gar nicht	ein wenig	etwas	sehr	äußerst

14. Ich habe versucht mich selbst kennenzulernen:

o	o	o	o	o
gar nicht	ein wenig	etwas	sehr	äußerst

15. Ich habe über mich selbst nachgedacht:

o	o	o	o	o
gar nicht	ein wenig	etwas	sehr	äußerst

16. Ich hatte einen Tagtraum:

o	o	o	o	o
gar nicht	ein wenig	etwas	sehr	äußerst

17. Ich fühlte mich über meine Fähigkeiten bewusst:

o	o	o	o	o
gar nicht	ein wenig	etwas	sehr	äußerst

18. Ich fühlte mich Selbstbewusst:

o	o	o	o	o
gar nicht	ein wenig	etwas	sehr	äußerst

19. Ich habe mir Gedanken darüber gemacht, was andere über mich denken:

o	o	o	o	o
gar nicht	ein wenig	etwas	sehr	äußerst

20. Ich war besorgt über den Eindruck, den ich machte:

o	o	o	o	o
gar nicht	ein wenig	etwas	sehr	äußerst

21. Ich habe diese Aufgabe gut gelöst:

o	o	o	o	o
gar nicht	ein wenig	etwas	sehr	äußerst

22. Im Allgemeinen hatte ich das Gefühl, die Dinge unter Kontrolle zu haben:

o	o	o	o	o
gar nicht	ein wenig	etwas	sehr	äußerst

23. Ich habe darüber nachgedacht, wie andere bei dieser Aufgabe abgeschnitten haben:

o	o	o	o	o
gar nicht	ein wenig	etwas	sehr	äußerst

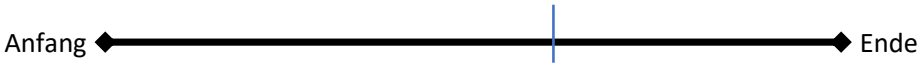
24. Ich habe darüber nachgedacht, wie es wäre, wenn mir gesagt werden würde, wie ich bei dieser Aufgabe abgeschnitten habe:

o	o	o	o	o
gar nicht	ein wenig	etwas	sehr	äußerst

Teil 3:

Bitte schätzen Sie keinen, einen oder mehrere Zeitpunkte durch eine vertikale Line auf dem Zeitstrahl ein:

Beispiel:



25. Ich habe meine Konzentration verloren: □ nein, wenn ja wann?



26. Ich war unaufmerksam: □ nein, wenn ja wann?



27. Ich war abgelenkt: □ nein, wenn ja wann?



28. Ich war müde: □ nein, wenn ja wann?



29. Ich hatte Sekundenschlaf: □ nein, wenn ja wann?



30. Ich hatte einen Tagtraum: □ nein, wenn ja wann?



Auswertung der Selbsteinschätzung:

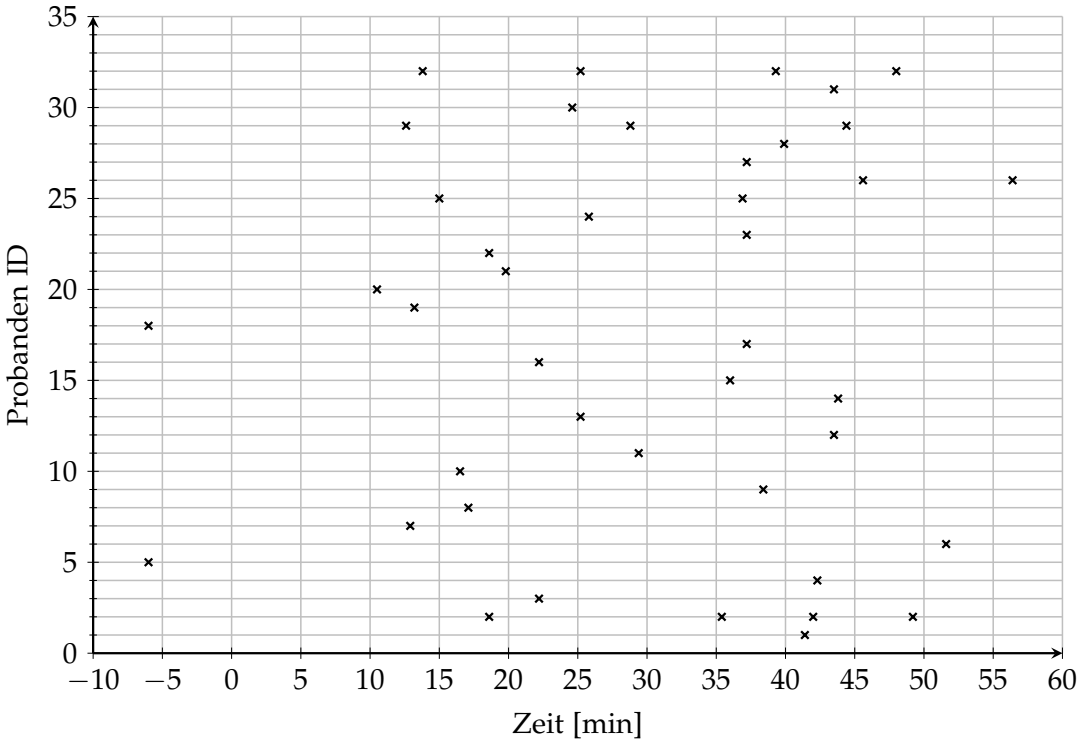


Abbildung A.4.: Selbsteinschätzung der Probanden zu Konzentrationsverlust

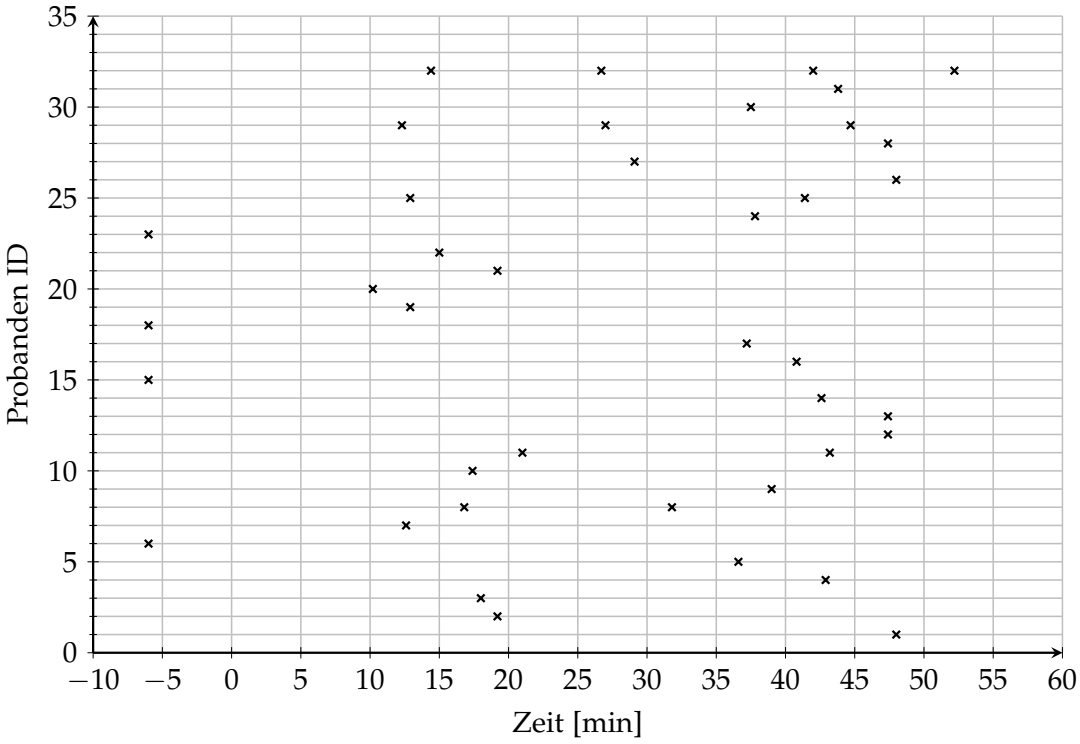


Abbildung A.5.: Selbsteinschätzung der Probanden zu Unaufmerksamkeit

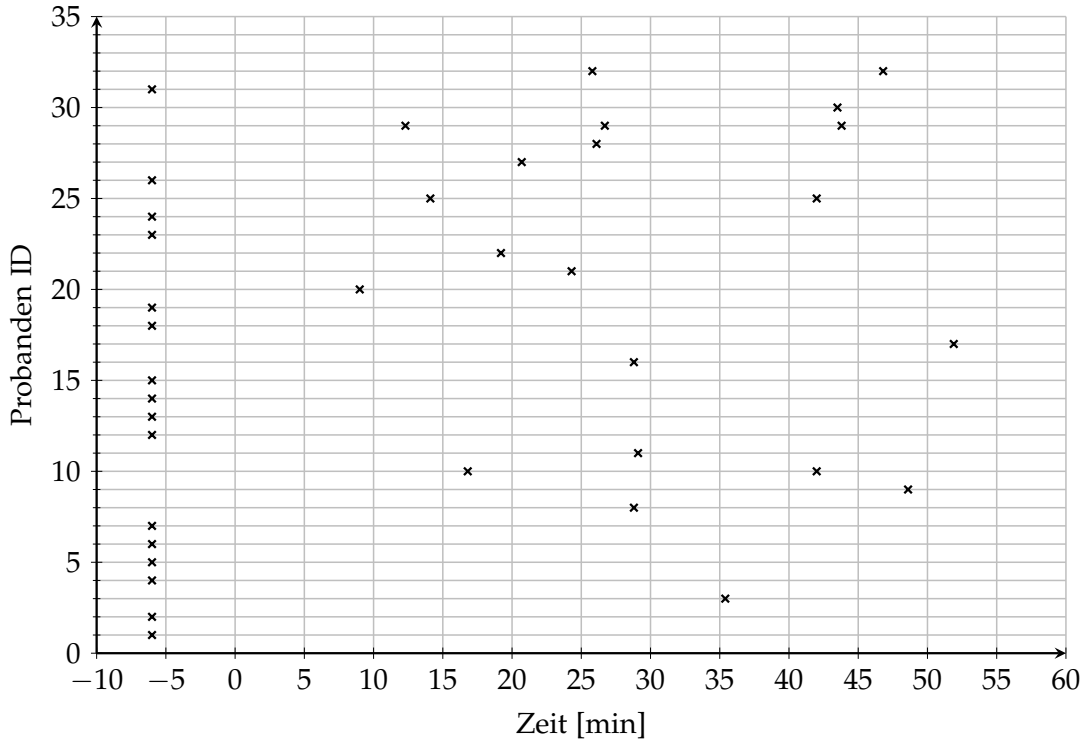


Abbildung A.6.: Selbsteinschätzung der Probanden zu Ablenkungen

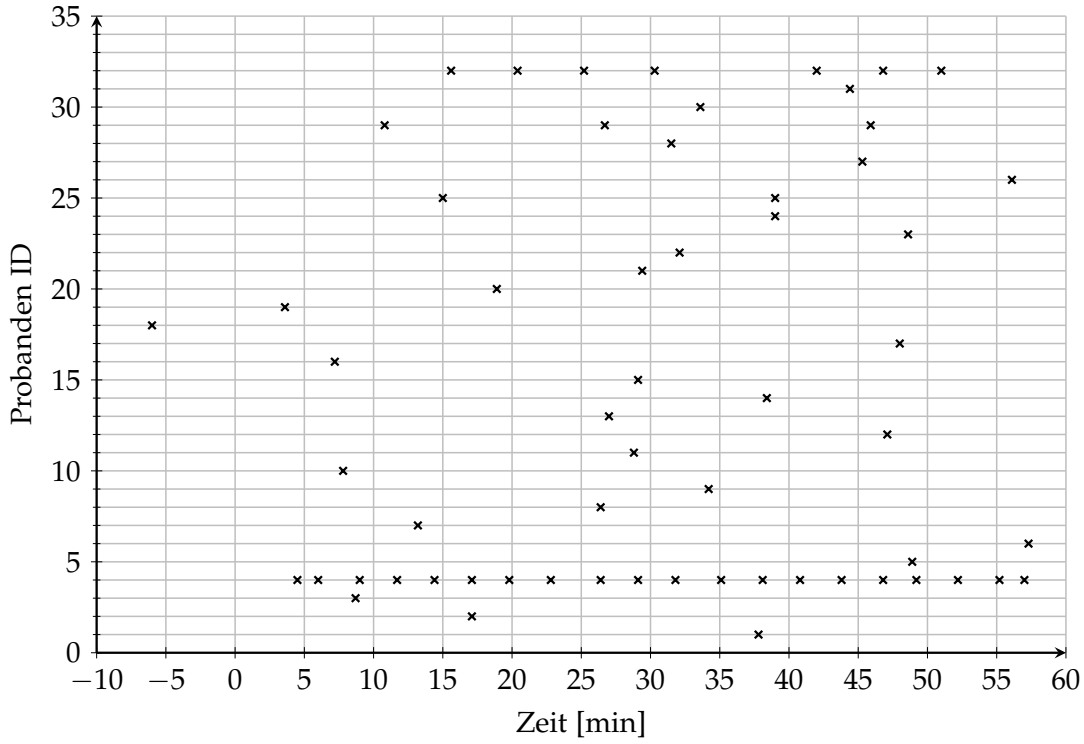


Abbildung A.7.: Selbsteinschätzung der Probanden zu Müdigkeit

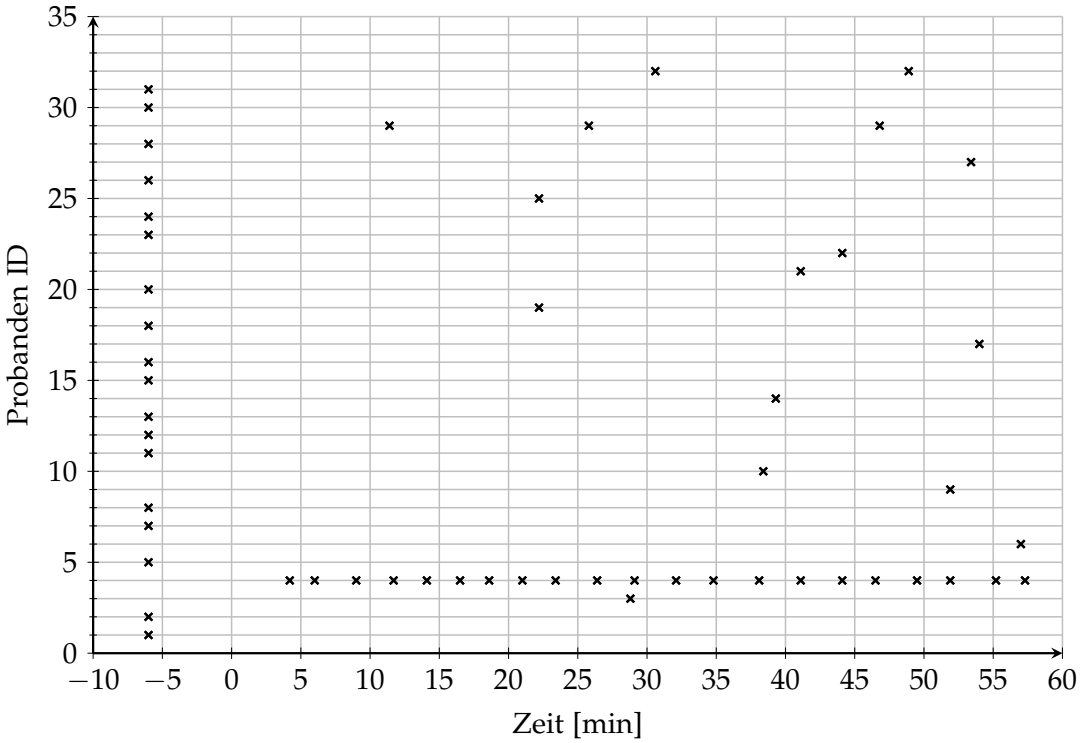


Abbildung A.8.: Selbsteinschätzung der Probanden zum Sekundenschlaf

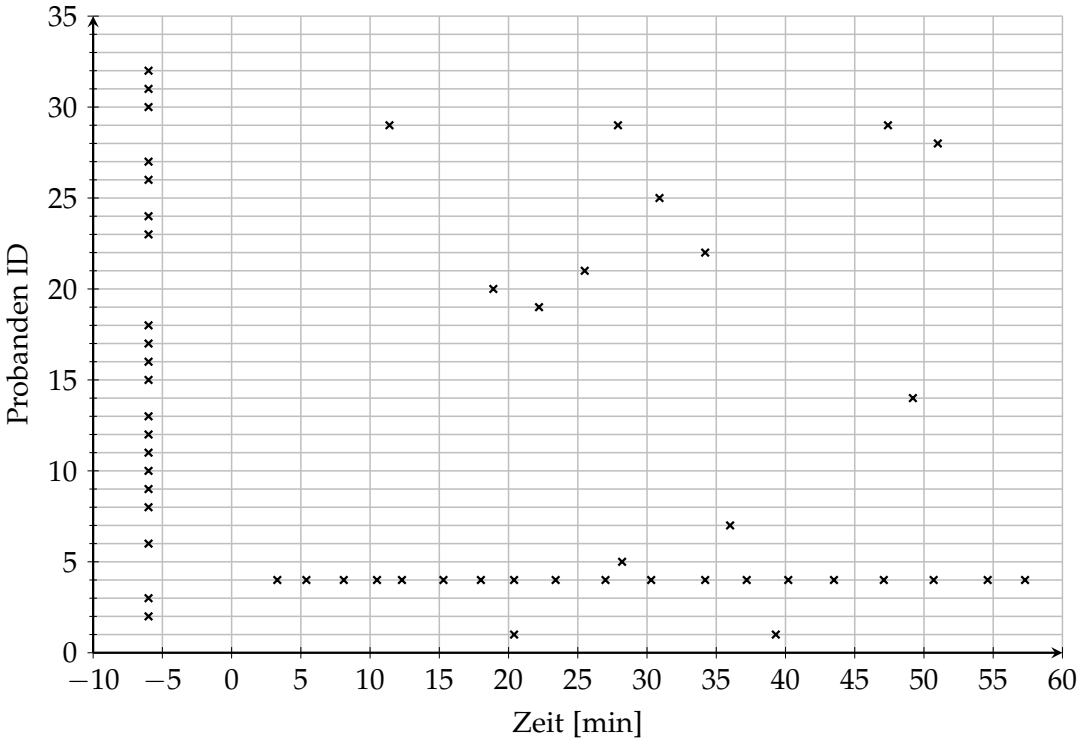


Abbildung A.9.: Selbsteinschätzung der Probanden zu Tagträumen

Literatur

[ÅG90] T. Åkerstedt und M. Gillberg. „Subjective and Objective Sleepiness in the Active Individual“. In: *International Journal of Neuroscience* 52.1-2 (1990), S. 29–37.

[Ake+14] T. Akerstedt, A. Anund, J. Axelsson und G. Kecklund. „Subjective sleepiness is a sensitive indicator of insufficient sleep and impaired waking function“. In: *Journal of sleep research* 23.3 (2014), S. 240–252.

[Arm+16] N. Armanfard, M. Komeili, J. P. Reilly und L. Pino. „Vigilance lapse identification using sparse EEG electrode arrays“. In: *IEEE Canadian Conference on Electrical and Computer Engineering (CCECE)* (2016). URL: https://ieeexplore.ieee.org/ielx7/7589833/7726593/07726846.pdf?tp=&arnumber=7726846&isnumber=7726593.

[Bau+01] S. D. Baulk, L. A. Reyner und J. A. Horne. „Driver sleepiness–evaluation of reaction time measurement as a secondary task“. In: *Sleep* 24.6 (2001), S. 695–698.

[Bib+19] D. Bibbo, M. Carli, S. Conforto und F. Battisti. „A Sitting Posture Monitoring Instrument to Assess Different Levels of Cognitive Engagement“. In: *Sensors (Basel, Switzerland)* 19.3 (2019).

[Cab+16] C. D. D. Cabrall, R. Happee und J. de Winter. „From Mackworth's clock to the open road: A literature review on driver vigilance task operationalization“. In: *Transportation Research Part F: Traffic Psychology and Behaviour* 40 (2016), S. 169–189.

[Chu+14] C.-H. Chuang, L.-W. Ko, T.-P. Jung und C.-T. Lin. „Kinesthesia in a sustained-attention driving task“. In: *NeuroImage* 91 (2014), S. 187–202.

[EF14] M. R. Esco und A. A. Flatt. „Ultra-Short-Term Heart Rate Variability Indexes at Rest and Post-Exercise in Athletes: Evaluating the Agreement with Accepted Recommendations“. In: *Journal of Sports Science & Medicine* 13.3 (2014), S. 535–541.

[Elm02] W. Elmenreich. „An introduction to sensor fusion“. In: *Vienna University of Technology, Austria* 502 (2002), S. 1–28.

[Fel+18a] A. Feldhütter, A. Feierle, L. Kalb und K. Bengler. „A New Approach for a Real-Time Non-Invasive Fatigue Assessment System for Automated Driving: Proceedings of the Human Factors and Ergonomics Society Annual Meeting“. In: *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 62 (1 2018), S. 1669–1673.

[Fel+18b] A. Feldhütter, T. Hecht und K. Bengler. „Fahrerspezifische Aspekte beim hochautomatisierten Fahren: Schlussbericht zum Forschungsprojekt FE 82.0628/2015: Schlussbericht zum Forschungsprojekt FE 82.0628/2015 im Auftrag der Bundesanstalt für Straßenwesen“. In: (2018).

[Gad+20] K. Gadzicki, R. Khamsehashari und C. Zetzsche. „Early vs Late Fusion in Multimodal Convolutional Neural Networks“. In: *2020 IEEE 23rd International Conference on Information Fusion (FUSION)*. IEEE, 2020.

[Gil+19] R. Gilgen-Ammann, T. Schweizer und T. Wyss. „RR interval signal quality of a heart rate monitor and an ECG Holter at rest and during exercise“. In: *European journal of applied physiology* 119.7 (2019), S. 1525–1532.

[Haa07] A. Haarmann. „Der Einsatz psychophysiologischer Variablen bei der adaptiven Automatisierung am Beispiel einer Flugsimulationsaufgabe“. Dissertation. Wuppertal: Bergische Universität Wuppertal, 2007. URL: http://nbn-resolving.de/urn/resolver.pl?urn=urn%3Anbn%3Ade%3Ahbz%3A468-20070925.

[Hec+19] T. Hecht, A. Feldhütter, J. Radlmayr, Y. Nakano, Y. Miki, C. Henle und K. Bengler. „A Review of Driver State Monitoring Systems in the Context of Automated Driving“. In: *Proceedings of the 20th Congress of the International Ergonomics Association (IEA 2018)*. Bd. 823. Advances in intelligent systems and computing. Cham: Springer International Publishing, 2019, S. 398–408.

[HN15] W. S. Helton und K. Näswall. „Short Stress State Questionnaire“. In: *European Journal of Psychological Assessment* 31.1 (2015), S. 20–30.

[Hor+89] K. Hornik, M. Stinchcombe und H. White. „Multilayer feedforward networks are universal approximators“. In: *Neural Networks* 2.5 (1989), S. 359–366.

[HS97] S. Hochreiter und J. Schmidhuber. „Long short-term memory“. In: *Neural computation* 9.8 (1997), S. 1735–1780.

[Int15] Interaxon. *MuseIO | Available Data*. 2015. URL: https://web.archive.org/web/20181105231756/http://developer.choosemuse.com/tools/available-data#Absolute_Band_Powers.

[Jab+18] R. Jabbar, K. Al-Khalifa, M. Kharbeche, W. Alhajyaseen, M. Jafari und S. Jiang. „Real-time Driver Drowsiness Detection for Android Application Using Deep Neural Networks Techniques“. In: *Procedia Computer Science* 130 (2018), S. 400–407.

[Jai+16] A. Jain, A. Singh, H. S. Koppula, S. Soh und A. Saxena. „Recurrent Neural Networks for driver activity anticipation via sensory-fusion architecture“. In: *2016 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2016.

[Jin+17a] Z. Jin, J. Cao, H. Guo, Y. Zhang und J. Luo. „Multimodal Fusion with Recurrent Neural Networks for Rumor Detection on Microblogs“. In: *Proceedings of the 25th ACM international conference on Multimedia*. New York, NY, USA: ACM, 2017.

[Jin+17b] L. Jing, T. Wang, M. Zhao und P. Wang. „An Adaptive Multi-Sensor Data Fusion Method Based on Deep Convolutional Neural Networks for Fault Diagnosis of Planetary Gearbox“. In: *Sensors (Basel, Switzerland)* 17.2 (2017), S. 414. URL: https://www.mdpi.com/1424-8220/17/2/414.

[Jür+18] T. Jürgensohn, S. Böhm, D. Gardas und T. Stephani. *Entwicklung der Fahr- und Verkehrskompetenz mit zunehmender Fahrerfahrung*. Berichte der Bundesanstalt für Strassenwesen Mensch und Sicherheit. Bremen: Carl Ed. Schünemann KG, 2018.

[Kör+15] M. Körber, A. Cingel, M. Zimmermann und K. Bengler. „Vigilance Decrement and Passive Fatigue Caused by Monotony in Automated Driving“. In: *Procedia Manufacturing* 3 (2015), S. 2403–2409.

[Kra+09] J. Krajewski, D. Sommer, U. Trutschel, D. Edwards und M. Golz. „Steering wheel behavior based estimation of fatigue“. In: *Proceedings of the Fifth International Driving Symposium on Human Factors in Driver Assessment* (2009), S. 118–124.

[KS67] J. Kalsbeek und R. N. Sykes. „Objective measurement of mental load“. In: *Acta Psychologica* 27 (1967), S. 253–261.

[LC13] G. Li und W.-Y. Chung. „Detection of Driver Drowsiness Using Wavelet Analysis of Heart Rate Variability and a Support Vector Machine Classifier“. In: *Sensors* 13 (2013), S. 16494–16511.

[Lee+17] J. Lee, J. Kim und M. Shin. „Correlation Analysis between Electrocardiography (ECG) and Photoplethysmogram (PPG) Data for Driver's Drowsiness Detection Using Noise Replacement Method“. In: *Procedia Computer Science* 116 (2017), S. 421–426.

[Leo+18] S. Leonhardt, L. Leicht und D. Teichmann. „Unobtrusive Vital Sign Monitoring in Automotive Environments—A Review“. In: *Sensors* 18.9 (2018).

[Mac48] N. H. Mackworth. „The breakdown of vigilance during prolonged visual search“. In: *Quarterly Journal of Experimental Psychology* (1948), S. 6–21.

[Mar+15] G. Marquart, C. Cabrall und J. de Winter. „Review of Eye-related Measures of Drivers' Mental Workload“. In: *Procedia Manufacturing* 3 (2015), S. 2854–2861.

[Mar+18] C. Marberger, H. Mielenz, F. Naujoks, J. Radlmayr, K. Bengler und B. Wandtner. „Understanding and Applying the Concept of “Driver Availability” in Automated Driving“. In: *Advances in Human Aspects of Transportation*. Hrsg. von N. A. Stanton. Bd. 597. Advances in intelligent

systems and computing. Cham: Springer International Publishing, 2018, S. 595–605.

[MP43] W. S. McCulloch und W. Pitts. „A logical calculus of the ideas immanent in nervous activity“. In: *The Bulletin of Mathematical Biophysics* 5.4 (1943), S. 115–133.

[Nau+17] C. J. de Naurois, C. Bourdin, A. Stratulat, E. Diaz und J.-L. Vercher. „Detection and prediction of driver drowsiness using artificial neural network models“. In: *Accident Analysis and Prevention* (2017).

[Nov19] D. Novak, Hrsg. *Chapter Five - Biomechatronic Applications of Brain-Computer Interfaces*. Handbook of Biomechatronics. Elsevier, 2019.

[Pop+95] A. T. Pope, E. H. Bogart und D. S. Bartolome. „Biocybernetic system evaluates indices of operator engagement in automated task“. In: *Biological Psychology* 40.1-2 (1995), S. 187–195.

[Rog+02] J. Rogé, T. Pébayle, L. Kiehn und A. Muzet. „Alteration of the useful visual field as a function of state of vigilance in simulated car driving“. In: *Transportation Research Part F: Traffic Psychology and Behaviour* 5.3 (2002), S. 189–200.

[Ros58] F. Rosenblatt. „The perceptron: a probabilistic model for information storage and organization in the brain“. In: *Psychological review* 65.6 (1958), S. 386–408.

[Sah+12a] A. Sahayadhas, K. Sundaraj und M. Murugappan. „Detecting Driver Drowsiness Based on Sensors: A Review“. In: *Sensors* 12 (2012), S. 16937–16953.

[Sah+12b] A. Sahayadhas, K. Sundaraj und M. Murugappan. „Detecting driver drowsiness based on sensors: a review“. In: *Sensors* 12.12 (2012), S. 16937–16953.

[Sch+18] E. Schires, P. Georgiou und T. S. Lande. „Vital Sign Monitoring Through the Back Using an UWB Impulse Radar With Body Coupled Antennas“. In: *IEEE Transactions on Biomedical Circuits and Systems* 12.2 (2018), S. 292–302.

[SG10] D. Sommer und M. Golz. „Evaluation of PERCLOS based current fatigue monitoring technologies“. In: *2010 Annual International Conference of the IEEE Engineering in Medicine and Biology*. IEEE. 2010, S. 4456–4459.

[SG17] F. Shaffer und J. P. Ginsberg. „An Overview of Heart Rate Variability Metrics and Norms“. In: *Frontiers in Public Health* 5.258 (2017).

[Sib+17] S. Sibi, S. Baiters, B. Mok, M. Steiner und W. Ju. „Assessing driver cortical activity under varying levels of automation with functional near infrared spectroscopy“. In: *2017 IEEE Intelligent Vehicles Symposium (IV)*. IEEE, 2017, S. 1509–1516.

[Sta+19] J. Stapel, F. A. Mullakkal-Babu und R. Happee. „Automated driving reduces perceived workload, but monitoring causes higher cognitive load than manual driving“. In: *Transportation Research Part F: Traffic Psychology and Behaviour* 60 (2019), S. 590–605.

[Tat+18] S. Tateno, X. Guan, R. Cao und Z. Qu. „Development of Drowsiness Detection System Based on Respiration Changes Using Heart Rate Monitoring“. In: *2018 57th Annual Conference of the Society of Instrument and Control Engineers of Japan (SICE)*. IEEE, 2018, S. 1664–1669.

[WE94] W. W. Wierwille und L. A. Ellsworth. „Evaluation of driver drowsiness by trained raters“. In: *Accident Analysis & Prevention* 26.5 (1994), S. 571–581.

[Wee+98] H.-G. Weeß, R. Lund, C. Gresele, W. Böhning, C. Sauter und R. Steinberg. „Vigilanz, Einschlafneigung, Daueraufmerksamkeit, Müdigkeit, Schläfrigkeit“. In: *Somnologie - Schlafforschung und Schlafmedizin* 2.1 (1998), S. 32–41.

[Wel38] B. L. Welch. „The Significance of the Difference Between Two Means when the Population Variances are Unequal“. In: *Biometrika* 29.3/4 (1938), S. 350.

[WF13] D. Wentura und C. Frings. *Kognitive Psychologie*. Wiesbaden: Springer Fachmedien Wiesbaden, 2013.

[Wu+12] H.-Y. Wu, M. Rubinstein, E. Shih, J. Guttag, F. Durand und W. Freeman. „Eulerian Video Magnification for Revealing Subtle Changes in the World“. In: *ACM Transactions on Graphics* 31.5 (2012), S. 1–8.

[Xia+17] Xiaoliang Zhang, Jiali Li, Yugang Liu, Zutao Zhang, Zhuojun Wang, Dianyuan Luo, Xiang Zhou, Miankuan Zhu, Waleed Salman und Guangdi Hu and Chunbai Wang. „Design of a Fatigue Detection System for High-Speed Trains Based on Driver Vigilance Using a Wireless Wearable EEG“. In: *Sensors* 2017 (2017).

[Xia+19] Z. Xiao, Z. Hu, L. Geng, F. Zhang, J. Wu und Y. Li. „Fatigue driving recognition network: fatigue driving recognition via convolutional neural network and long short–term memory units“. In: *IET Intelligent Transport Systems* 13.9 (2019), S. 1410–1416.

[Z C+19] Z. Cao, G. Hidalgo Martinez, T. Simon, S. Wei und Y. A. Sheikh. „OpenPose: Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields“. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2019).

[Zha+17a] D.-q. Zhang, W.-j. Chen und H.-z. Yang. „Attention State Related EEG Spectrum and Pupil Size in Vigilance Task“. In: *International Conference on Mechanical Control and Automation* (2017), S. 231–240. URL: `http://www.dpi-proceedings.com/index.php/dtetr/article/viewFile/12372/11908`.

[Zha+17b] X. Zhang, J. Li, Y. Liu, Z. Zhang, Z. Wang, D. Luo, X. Zhou, M. Zhu, W. Salman, G. Hu und et al. „Design of a Fatigue Detection System for High-Speed Trains Based on Driver Vigilance Using a Wireless Wearable EEG“. In: *Sensors* 17.3 (März 2017), S. 486. URL: http://dx.doi.org/10.3390/s17030486.

[Zha06] J. Zhang. „Improved on-line process fault diagnosis through information fusion in multiple neural networks“. In: *Computers & Chemical Engineering* 30.3 (2006), S. 558–571.

[Zsc02] S. Zschocke. *Klinische Elektroenzephalographie*. Springer eBook Collection Medicine. Berlin, Heidelberg und s.l.: Springer Berlin Heidelberg, 2002.

Relevante wissenschaftliche Publikationen

Liste der Publikationen, die im Zusammanhang mit dieser Dissertation stehen:

A. Hugenroth, A. Warnecke und T. Bertram. „Untersuchung und Klassifikation der Fahreraufmerksamkeit bei längerer partiell automatisierter Fahrt“. In: *Forschung im Ingenieurwesen* 86.1 (2022), S. 49–63.

A. Hugenroth, A. Warnecke und T. Bertram. „Mechatronisches System für die modellbasierte Datenfusion zur Ermittlung des psychologischen Fahrerzustands“. In: *Digital-Fachtagung VDI MECHATRONIK 2021*. VDI. 2021, S. 146–151.

A. Hugenroth, F. Albers, M. Oeljeklaus und T. Bertram. „Object Classification in the Fixation of a Car Driver“. In: *AmE 2020-Automotive meets Electronics; 11. GMM-Symposium*. VDE. 2020, S. 1–6.

D. N. Khazar, A. Hugenroth und T. Bertram. „Position Classification and In-Vehicle Activity Detection Using Seat-Pressure-Sensor in Automated Driving“. In: *AmE 2022-Automotive meets Electronics; 13. GMM-Symposium*. VDE. 2022, S. 1–6.

Betreute Abschlussarbeiten

L.-G. Negoita. „Driver monitoring based on seat-pressure-sensor“. Masterarbeit. 2021.

J. A. J. Merida. „Evaluation eines Sitzdrucksystems zur Sitzposen-Erkennung mit neuronalen Netzwerken durch Kamerabilder“. Bachelorarbeit. 2020.

T. Bednorz. „Entwicklung eines Sensor-konzepts zur Fahrerüberwachung im Fahrzeuginnenraum“. Bachelorarbeit. 2019.