Driver Mental States Detection during Highly Automated Driving by Decoding Brain Signals

vorgelegt von
M. Sc.
Xixie Zhang

an der Fakultät V – Verkehrs- und Maschinensysteme
der Technischen Universität Berlin
zur Erlangung des akademischen Grades

Doktorin der Naturwissenschaften
– Dr.rer.nat. –

genehmigte Dissertation

Promotionsausschuss:

Vorsitzender: Prof. Dr.-Ing. Henning Jürgen Meyer
Gutachter: Prof. Dr.-Ing. Matthias Rötting
Gutachter: Prof. Dr. Bao-Liang Lu

Tag der wissenschaftlichen Aussprache: 23.02.2021

Berlin 2021

Acknowledgement

Foremost, I am grateful to my advisor Prof. Dr.-Ing. Matthias Rötting for his continuous support of my Ph.D. study and research, for his patience, motivation, enthusiasm, and immense knowledge. His guidance helped me in all the time of research and writing of this thesis. I could not have imagined having a better advisor and mentor for my Ph.D. study.

I would also like to thank Prof. Dr. Baoliang Lu and his team at Shanghai Jiao Tong University for their support in the second and third experiment of this thesis. Without the opportunity to conduct my experiment at the BCMI Laboratory and their kind help I would hardly accomplish my goals.

I would also like to express my gratitude to Prof. Dr.-Ing. Henning Jürgen Meyer for chairing the defense committee.

My sincere thanks also go to all of my colleagues at Chair of Human-Machine-Systems for their kind help in my Ph.D. study. In particular, I would like to thank Zhe Zeng, Xiaohan Li and Yuan Liu for cheering me up when needed and sharing their experience with me. Furthermore, a special thank goes to Dr. Thorsten Zander for enlightening me at the beginning of my research and supporting me in the first experiment.

Last but not least, I am thankful to my husband, Mr. Keshen Bao, for supporting me spiritually throughout my life. I would also like to thank my family: my parents Jianzhong Zhang and Dongmei Xu, for their patience, support and endless love, and my parents-in-law Xinhe Bao and Guoyu Wang for their thoughtful care and guidance.

List of contents

Summary **9**

Zusammenfassung **11**

1. Introduction **13**

1.1 Background 13

1.2 Aims of the thesis 14

1.3 Structure of the thesis 15

2. Autonomous driving **17**

2.1 Introduction to autonomous driving 17

2.2 Classification of autonomous driving levels 18

2.3 Development of autonomous driving 21

2.4 Take-over control in highly automated driving 24

3. Driver states monitoring in HAD **27**

3.1 Importance of driver monitoring in HAD 27

3.2 Different mental states 29

3.3 Aspects investigated in this thesis 30

4. Mental states investigated in this thesis **33**

4.1 Mental workload 33

4.1.1 Definition 33

4.1.2 Measurements 34

4.1.2.1 Performance measures 34

4.1.2.2 Subjective measures 35

4.1.2.3 Physiological measures 36

4.2 Vigilance 38

4.2.1 Definition 38

4.2.2 Measurements 39

5. Electroencephalography **43**

5.1 Definition 43

5.2 Measuring mental workload with EEG 49

5.3 Measuring vigilance with EEG 51

6. EEG analysis methods **53**

6.1 Signal processing procedure 53

6.2 Preprocessing and feature extraction 54

6.2.1 Preprocessing ... 54
6.2.2 Feature extraction ... 56
6.2.3 Common Spatial Patterns ... 58
6.3 Classification ... 60
6.3.1 Different classifiers ... 60
6.3.2 Linear Discriminant Analysis ... 61
6.4 Method applied in this work ... 63
7. Experiment 1 – Validation of task-independent workload classifier with EEG ... 65
7.1 Introduction and hypothesis ... 65
7.2 Introduction of the tasks ... 65
7.3 Pre-study: Manipulating workload of different tasks ... 67
7.4 Methods ... 69
7.4.1 Participants ... 69
7.4.2 Experimental design ... 69
7.4.3 Procedure ... 69
7.4.4 Experiment apparatus ... 70
7.4.5 Data analysis ... 70
7.5 Results ... 71
7.5.1 Subjective measures ... 71
7.5.2 Performance of task-independent workload classifier ... 72
7.5.3 CSP pattern analysis ... 77
7.6 Discussion ... 77
7.7 Summary ... 79
8. Experiment 2 – Detection of driver's mental workload and vigilance with EEG in highly automated driving ... 81
8.1 Introduction and hypothesis ... 81
8.2 Introduction of the tasks ... 82
8.2.1 Primary task ... 82
8.2.2 Secondary task ... 83
8.3 Pre-study: Manipulating workload of different non-driving tasks and vigilance under different workload conditions ... 83
9. Experiment 2a – Detection of driver's mental workload with EEG ... 89
9.1 Methods ... 89
9.1.1 Participants ... 89

9.1.2 Experimental design 89
9.1.3 Procedure 90
9.1.4 Experiment apparatus 91
9.1.5 Data analysis 94
9.2 Results 95
9.2.1 Subjective measures 95
9.2.2 Classification accuracy of direct training on non-driving tasks 99
9.2.3 Performance of task-independent workload classifier 102
9.3 Discussion 105
9.3.1 Subjective measures 105
9.3.2 Direct training classification accuracy 105
9.3.3 Task-independent workload classifier 106
9.4 Summary 107
10. Experiment 2b – Vigilance detection with EEG 109
10.1 Methods 109
10.1.1 Participants 109
10.1.2 Procedure 109
10.1.3 Data acquisition 111
10.1.4 Data analysis 111
10.2 Results 112
10.2.1 ERP analysis and classification accuracy 112
10.2.2 Performance 116
10.3 Discussion 120
10.4 Summary 122
11. Experiment 3 – Online detection of driver's mental workload with EEG in highly automated driving and its impact on take-over control 123
11.1 Introduction and hypothesis 123
11.2 Methods 124
11.2.1 Participants 124
11.2.2 Experimental design 124
11.2.3 Procedure 125
11.2.4 Experiment apparatus 127
11.2.5 Data analysis 129
11.3 Results 129

11.3.1 Subjective workload measure 129
11.3.2 Classification results 130
11.3.3 Take-over time 132
11.3.4 Feedback ratings 134
11.4 Discussion 136
11.5 Summary 137
12. Conclusions and outlook 139
12.1 Conclusions 139
12.2 Outlook 140
References 143
List of figures 163
List of tables 167
List of abbreviations 169
Appendix 171

Summary

In contrast to manual driving, highly automated driving may relieve the driver from active driving and observing activities. However, undertaking different non-driving tasks while highly automated driving could lead to variations in driver's mental workload and vigilance levels. This has great influence on driver's ability and performance to take over control of the vehicle when requested. Therefore, the overall goal of the work is to investigate methods to measure driver's mental workload and vigilance passively (by decoding brain signals) as that potentially improves driver's ability for take-over request in urgent situations. To achieve this, three separate studies were carried out.

The first study was aimed at validating a task-independent workload classifier with EEG in well-controlled laboratory. In this study, a new methodology was investigated and verified to predict workload levels in different tasks. More specifically, an individual classifier was trained on a standardized workload-inducing task in the calibration phase, and then applied to data from five other cognitive tasks each inducing two workload levels. Results showed that a direct transfer across tasks was not optimal to separate high and low workload levels with acceptable accuracy, however, significant differences were found in the classifier outputs between two conditions of five tasks after applying this task-independent classifier. It was implied to look at the classifier output on a continuous scale as an indication of the actual workload level, rather than just take the binary classification accuracy into consideration. Moreover, it was possible to calibrate a single classifier on a standardized workload-inducing task and apply this classifier to other tasks to obtain meaningful results on workload levels. This greatly simplifies the calibration process of different tasks and enables an easy implementation of real-time and continuous workload detection with EEG for future real-world applications.

Based on the results from the first study, the task-independent workload classifier was then applied in the second study to classify workload levels of different non-driving tasks during highly automated driving. In this study, detection of vigilance with EEG and the impact of different workload levels on vigilance were also investigated with an auditory oddball task. It was proven that the simplified task-independent workload classifier was able to indicate mental workload of non-driving tasks with classifier outputs, which can significantly differentiate between classes. This task-independent workload classifier was demonstrated as an applicable tool in more realistic and practical scenarios. Moreover, oddball task results indicated the feasibility of detecting signal perception with EEG, thus inferring driver's vigilance level based on this. Besides, higher mental workload induced by non-driving tasks resulted in deteriorated vigilance to detect target stimuli, and this variation in target stimuli detection confirmed the influence of workload on driver's vigilance level.

Finally, on the basis of the feasibility of task-independent workload classifier, an online study of workload detection is conducted in the third study, which investigated the effect of real-time warning based on workload measurement on the reaction in take-over request. This study examined the impact of feedback appearance on driver's take-over performance in a simulated highly automated driving scenario. Results indicated that it was possible to identify driver's workload level in real time with the task-independent workload classifier. Furthermore, online feedback to overload condition was proven to be useful in improving driver's take-over performance. In addition, subjective ratings on feedback appearance also confirmed its positive effect.

From these three studies we could conclude that, it is possible to measure driver's mental workload and vigilance levels during highly automated driving by decoding brain signals, and this supports the improvement of driver's take-over performance. The outcomes of these studies could contribute to the design of driver-vehicle-interaction in highly autonomous driving by identifying driver's mental states during mode transition, and thus enhance the safety and user experience of drivers.

03.03.2021

Zusammenfassung

Im Vergleich zum manuellen Fahren kann das hochautomatisiertes Fahren den Fahrer vom aktiven Fahren und Beobachten der Umgebung entlasten. Das Ausführen verschiedener Nebenaufgaben während des hochautomatisierten Fahrens kann jedoch zur Variation der mentalen Beanspruchung und der Vigilanz des Fahrers führen. Dies hat großen Einfluss auf die Fähigkeit und Leistung des Fahrers, auf Anfrage die Kontrolle über das Fahrzeug zu übernehmen. Daher besteht das Ziel dieser Arbeit vor allem darin, Methoden zur passiven Messung der mentalen Beanspruchung und Vigilanz des Fahrers (durch Analyse von Gehirnsignalen) zu untersuchen, da dies möglicherweise die Fähigkeit des Fahrers zur Übernahme in dringenden Situationen verbessert. Um dies zu erreichen, wurden drei separate Studien durchgeführt.

Die erste Studie zielte darauf ab, einen aufgabenunabhängigen Beanspruchung-Klassifikator mit EEG in einem gut kontrollierten Labor zu validieren. In dieser Studie wurde eine neue Methodik untersucht und verifiziert, um die mentale Beanspruchung in verschiedenen Aufgaben zu messen. Insbesondere wurde ein individueller Klassifikator in der Kalibrierungsphase auf eine standardisierte Beanspruchung-induzierende Aufgabe trainiert und dann auf Daten von fünf anderen kognitiven Aufgaben angewendet, die jeweils zwei Beanspruchungsstufen hatten. Die Ergebnisse zeigten, dass eine direkte Übertragung zwischen Aufgaben nicht optimal war, um hohe und niedrige Beanspruchungsstufen mit akzeptabler Genauigkeit zu trennen. Es gab jedoch signifikante Unterschiede in den Klassifizierungsausgaben zwischen zwei Bedingungen der fünf Aufgaben nach Anwendung dieses aufgabenunabhängigen Klassifikators. Es wurde impliziert, die Klassifizierungsausgabe auf einer kontinuierlichen Skala als Hinweis auf die tatsächliche Beanspruchung zu betrachten, anstatt nur die Genauigkeit der binären Klassifizierung zu berücksichtigen. Darüber hinaus war es möglich, einen individuellen Klassifikator für eine standardisierte Beanspruchung-induzierende Aufgabe zu kalibrieren und diesen Klassifikator auf andere Aufgaben anzuwenden, um aussagekräftige Ergebnisse für Beanspruchungsstufe zu erhalten. Dies vereinfacht den Kalibrierungsprozess für verschiedene Aufgaben erheblich und ermöglicht eine einfache Implementierung der echtzeitigen und kontinuierlichen Erkennung der mentalen Beanspruchung mit EEG für zukünftige reale Anwendungen.

Basierend auf den Ergebnissen der ersten Studie wurde dann in der zweiten Studie der aufgabenunabhängige Beanspruchung-Klassifikator angewendet, um die Beanspruchungsstufen verschiedener Nebenaufgaben während des hochautomatisierten Fahrens zu klassifizieren. In dieser Studie wurden die Messung der Vigilanz mit EEG und auch die Auswirkungen unterschiedlicher Beanspruchungsstufen auf die Vigilanz mit einer auditorischen Oddball-Aufgabe untersucht. Es wurde nachgewiesen, dass der vereinfachte aufgabenunabhängige Beanspruchung-Klassifikator in der Lage war, die mentale Beanspruchung mit

Klassifizierungsausgaben anzuzeigen, die signifikant zwischen Klassen unterscheiden können. Dieser aufgabenunabhängige Beanspruchung-Klassifikator wurde als anwendbares Werkzeug in realistischen und praktischen Szenarien demonstriert. Darüber hinaus zeigten die Ergebnisse der Oddball-Aufgabe, dass es möglich ist, die Signalwahrnehmung mit EEG zu erfassen und daraus auf der Grundlage der Vigilanz des Fahrers zu schließen. Außerdem führte eine höhere mentale Beanspruchung, die durch Nebenaufgaben verursacht wurde, zu einer verminderten Vigilanz bei der Erkennung von Zielreizen, und diese Variation bei der Erkennung von Zielreizen bestätigte den Einfluss der mentalen Beanspruchung auf das Wachsamkeitsniveau des Fahrers.

Auf der Grundlage der Machbarkeit eines aufgabenunabhängigen Beanspruchung-Klassifikators wird schließlich in der dritten Studie eine Online-Studie zur Erkennung der mentalen Beanspruchung durchgeführt, in der die Auswirkung einer Echtzeitwarnung, auf der Grundlage der Beanspruchungsklassifizierung, auf die Reaktion bei Übernahmeanfragen untersucht wurde. Diese Studie untersuchte die Auswirkungen des Auftretens der Rückmeldungen auf die Übernahmeleistung des Fahrers in einem simulierten hochautomatisierten Fahrszenario. Die Ergebnisse zeigten, dass es möglich war, die Beanspruchungsstufen des Fahrers in Echtzeit mit dem aufgabenunabhängigen Beanspruchung-Klassifikator zu identifizieren. Darüber hinaus haben sich die Online-Rückmeldungen zum Überlastungszustand als nützlich erwiesen, um die Übernahmeleistung des Fahrers zu verbessern. Außerdem bestätigten subjektive Bewertungen des Auftretens der Rückmeldungen auch seine positive Wirkung.

Aus diesen drei Studien konnten wir schließen, dass es möglich ist, die mentale Beanspruchung und Vigilanz des Fahrers während des hochautomatisierten Fahrens durch Analyse von Gehirnsignalen zu messen, und dadurch die Übernahmeleistung des Fahrers zu verbessern. Die Ergebnisse dieser Studien könnten zur Gestaltung der Fahrer-Fahrzeug-Interaktion beim hochautonomen Fahren beitragen, indem sie die mentalen Zustände des Fahrers beim Wechseln der Modus identifizieren und daher die Sicherheit und das Benutzererlebnis des Fahrers verbessern.

03.03.2021

Summary

In contrast to manual driving, highly automated driving may relieve the driver from active driving and observing activities. However, undertaking different non-driving tasks while highly automated driving could lead to variations in driver's mental workload and vigilance levels. This has great influence on driver's ability and performance to take over control of the vehicle when requested. Therefore, the overall goal of the work is to investigate methods to measure driver's mental workload and vigilance passively (by decoding brain signals) as that potentially improves driver's ability for take-over request in urgent situations. To achieve this, three separate studies were carried out.

The first study was aimed at validating a task-independent workload classifier with EEG in well-controlled laboratory. In this study, a new methodology was investigated and verified to predict workload levels in different tasks. More specifically, an individual classifier was trained on a standardized workload-inducing task in the calibration phase, and then applied to data from five other cognitive tasks each inducing two workload levels. Results showed that a direct transfer across tasks was not optimal to separate high and low workload levels with acceptable accuracy, however, significant differences were found in the classifier outputs between two conditions of five tasks after applying this task-independent classifier. It was implied to look at the classifier output on a continuous scale as an indication of the actual workload level, rather than just take the binary classification accuracy into consideration. Moreover, it was possible to calibrate a single classifier on a standardized workload-inducing task and apply this classifier to other tasks to obtain meaningful results on workload levels. This greatly simplifies the calibration process of different tasks and enables an easy implementation of real-time and continuous workload detection with EEG for future real-world applications.

Based on the results from the first study, the task-independent workload classifier was then applied in the second study to classify workload levels of different non-driving tasks during highly automated driving. In this study, detection of vigilance with EEG and the impact of different workload levels on vigilance were also investigated with an auditory oddball task. It was proven that the simplified task-independent workload classifier was able to indicate mental workload of non-driving tasks with classifier outputs, which can significantly differentiate between classes. This task-independent workload classifier was demonstrated as an applicable tool in more realistic and practical scenarios. Moreover, oddball task results indicated the feasibility of detecting signal perception with EEG, thus inferring driver's vigilance level based on this. Besides, higher mental workload induced by non-driving tasks resulted in deteriorated vigilance to detect target stimuli, and this variation in target stimuli detection confirmed the influence of workload on driver's vigilance level.

Finally, on the basis of the feasibility of task-independent workload classifier, an online study of workload detection is conducted in the third study, which investigated the effect of real-time warning based on workload measurement on the reaction in take-over request. This study examined the impact of feedback appearance on driver's take-over performance in a simulated highly automated driving scenario. Results indicated that it was possible to identify driver's workload level in real time with the task-independent workload classifier. Furthermore, online feedback to overload condition was proven to be useful in improving driver's take-over performance. In addition, subjective ratings on feedback appearance also confirmed its positive effect.

From these three studies we could conclude that, it is possible to measure driver's mental workload and vigilance levels during highly automated driving by decoding brain signals, and this supports the improvement of driver's take-over performance. The outcomes of these studies could contribute to the design of driver-vehicle-interaction in highly autonomous driving by identifying driver's mental states during mode transition, and thus enhance the safety and user experience of drivers.

Zusammenfassung

Im Vergleich zum manuellen Fahren kann das hochautomatisiertes Fahren den Fahrer vom aktiven Fahren und Beobachten der Umgebung entlasten. Das Ausführen verschiedener Nebenaufgaben während des hochautomatisierten Fahrens kann jedoch zur Variation der mentalen Beanspruchung und der Vigilanz des Fahrers führen. Dies hat großen Einfluss auf die Fähigkeit und Leistung des Fahrers, auf Anfrage die Kontrolle über das Fahrzeug zu übernehmen. Daher besteht das Ziel dieser Arbeit vor allem darin, Methoden zur passiven Messung der mentalen Beanspruchung und Vigilanz des Fahrers (durch Analyse von Gehirnsignalen) zu untersuchen, da dies möglicherweise die Fähigkeit des Fahrers zur Übernahme in dringenden Situationen verbessert. Um dies zu erreichen, wurden drei separate Studien durchgeführt.

Die erste Studie zielte darauf ab, einen aufgabenunabhängigen Beanspruchung-Klassifikator mit EEG in einem gut kontrollierten Labor zu validieren. In dieser Studie wurde eine neue Methodik untersucht und verifiziert, um die mentale Beanspruchung in verschiedenen Aufgaben zu messen. Insbesondere wurde ein individueller Klassifikator in der Kalibrierungsphase auf eine standardisierte Beanspruchung-induzierende Aufgabe trainiert und dann auf Daten von fünf anderen kognitiven Aufgaben angewendet, die jeweils zwei Beanspruchungsstufen hatten. Die Ergebnisse zeigten, dass eine direkte Übertragung zwischen Aufgaben nicht optimal war, um hohe und niedrige Beanspruchungsstufen mit akzeptabler Genauigkeit zu trennen. Es gab jedoch signifikante Unterschiede in den Klassifizierungsausgaben zwischen zwei Bedingungen der fünf Aufgaben nach Anwendung dieses aufgabenunabhängigen Klassifikators. Es wurde impliziert, die Klassifizierungsausgabe auf einer kontinuierlichen Skala als Hinweis auf die tatsächliche Beanspruchung zu betrachten, anstatt nur die Genauigkeit der binären Klassifizierung zu berücksichtigen. Darüber hinaus war es möglich, einen individuellen Klassifikator für eine standardisierte Beanspruchung-induzierende Aufgabe zu kalibrieren und diesen Klassifikator auf andere Aufgaben anzuwenden, um aussagekräftige Ergebnisse für Beanspruchungsstufe zu erhalten. Dies vereinfacht den Kalibrierungsprozess für verschiedene Aufgaben erheblich und ermöglicht eine einfache Implementierung der echtzeitigen und kontinuierlichen Erkennung der mentalen Beanspruchung mit EEG für zukünftige reale Anwendungen.

Basierend auf den Ergebnissen der ersten Studie wurde dann in der zweiten Studie der aufgabenunabhängige Beanspruchung-Klassifikator angewendet, um die Beanspruchungsstufen verschiedener Nebenaufgaben während des hochautomatisierten Fahrens zu klassifizieren. In dieser Studie wurden die Messung der Vigilanz mit EEG und auch die Auswirkungen unterschiedlicher Beanspruchungsstufen auf die Vigilanz mit einer auditorischen Oddball-Aufgabe untersucht. Es wurde nachgewiesen, dass der vereinfachte aufgabenunabhängige Beanspruchung-Klassifikator in der Lage war, die mentale Beanspruchung mit

Klassifizierungsausgaben anzuzeigen, die signifikant zwischen Klassen unterscheiden können. Dieser aufgabenunabhängige Beanspruchung-Klassifikator wurde als anwendbares Werkzeug in realistischen und praktischen Szenarien demonstriert. Darüber hinaus zeigten die Ergebnisse der Oddball-Aufgabe, dass es möglich ist, die Signalwahrnehmung mit EEG zu erfassen und daraus auf der Grundlage der Vigilanz des Fahrers zu schließen. Außerdem führte eine höhere mentale Beanspruchung, die durch Nebenaufgaben verursacht wurde, zu einer verminderten Vigilanz bei der Erkennung von Zielreizen, und diese Variation bei der Erkennung von Zielreizen bestätigte den Einfluss der mentalen Beanspruchung auf das Wachsamkeitsniveau des Fahrers.

Auf der Grundlage der Machbarkeit eines aufgabenunabhängigen Beanspruchung-Klassifikators wird schließlich in der dritten Studie eine Online-Studie zur Erkennung der mentalen Beanspruchung durchgeführt, in der die Auswirkung einer Echtzeitwarnung, auf der Grundlage der Beanspruchungsklassifizierung, auf die Reaktion bei Übernahmeanfragen untersucht wurde. Diese Studie untersuchte die Auswirkungen des Auftretens der Rückmeldungen auf die Übernahmeleistung des Fahrers in einem simulierten hochautomatisierten Fahrszenario. Die Ergebnisse zeigten, dass es möglich war, die Beanspruchungsstufen des Fahrers in Echtzeit mit dem aufgabenunabhängigen Beanspruchung-Klassifikator zu identifizieren. Darüber hinaus haben sich die Online-Rückmeldungen zum Überlastungszustand als nützlich erwiesen, um die Übernahmeleistung des Fahrers zu verbessern. Außerdem bestätigten subjektive Bewertungen des Auftretens der Rückmeldungen auch seine positive Wirkung.

Aus diesen drei Studien konnten wir schließen, dass es möglich ist, die mentale Beanspruchung und Vigilanz des Fahrers während des hochautomatisierten Fahrens durch Analyse von Gehirnsignalen zu messen, und dadurch die Übernahmeleistung des Fahrers zu verbessern. Die Ergebnisse dieser Studien könnten zur Gestaltung der Fahrer-Fahrzeug-Interaktion beim hochautonomen Fahren beitragen, indem sie die mentalen Zustände des Fahrers beim Wechseln der Modus identifizieren und daher die Sicherheit und das Benutzererlebnis des Fahrers verbessern.

1. Introduction

1.1 Background

Compared to the cars one century ago, there has been a vast development in automobile technology. Due to the introduction of various safety-related technologies, it is evident that driving a car has become safer over the past few decades. The number of people killed in traffic in Germany decreased from 20.3 people in 1955 to 0.6 people in 2016 per 10,000 vehicles (Statistisches Bundesamt, 2017, p. 60). With the implementation of technologies such as Anti-lock Brake System (ABS) and Adaptive Cruise Control (ACC), drivers get more and more assistance to lower the impact of a crash or even to avoid critical situations.

However, still many people die of traffic accidents because of inattention or fatigue or other factors. One aim of introducing automated driving is to reduce the number of accidents and the damage in accidents. Automated driving or even Advanced Driver Assistance Systems (ADAS) are able to greatly decrease the risk of accidents. Chira-Chavala and Yoo (1994) calculated a potential reduction in the number of accidents by an Adaptive Cruise Control as 7.5%. With the use of Blind Spot Detection, Lane Departure Warning, and Collision Warning systems, there will be a potential reduction of 24% in the number of accidents (Harper, et al., 2016). Apart from the figures mentioned, more automated trucks and passenger cars could potentially prevent over two million accidents in Germany alone. In addition, automated driving is claimed to contribute to reducing vehicle consumption, improving traffic efficiency and driver comfort due to its reduced load (Brookhuis et al., 2001; Rhede, 2017).

On the way to driverless driving, a step-by-step introduction of automated systems with the takeover of longitudinal and lateral control has already taken place. Until driverless driving is fully implemented, it is likely that automated driving will be introduced in further intermediate stages and the driver will continue to be used to drive the vehicle. Compared to manual driving, the driver's task in these intermediate stages in the vehicle changes from an active driver to an observer of the performance of a system. Among the various levels of automation, in particular highly automated driving represents a major change in the task for the driver. Highly automated driving, following the definition of the *Bundesanstalt für Straßenwesen* (Gasser, et al., 2012), constitutes a system taking over the complete driving task for a specific period of time. When a system of this type is active, the driver is released from all driving and observing activities. The automation of driving can be a great relief, especially for driving phases in which vehicle control is perceived by the driver as a burden. Nevertheless, the driver must remain available during the entire journey and, if requested by the system, take over the driving task within a certain time. This new system changes the entire driving style.

During highly automated driving, drivers are potentially free to engage in non-driving tasks. This out-of-the-loop-behavior may lead to a contingently serious deterioration of situation awareness caused by a shift in driver attention from the vehicle status and surrounding traffic situation to the non-driving tasks. In such situation, the driver must be able to detect system boundaries that require a take-over within sufficient time for a safe and comfortable reaction. The driver should remain ready to respond to the execution of the driving at any time. The takeover and safety of the driver and the environment highly depend on his/her ability to take over. The task demand of these non-driving tasks may impact driver's workload niveau and vigilance level, thus influencing his/her readiness and ability to regain control of the vehicle. This raises the question of how the change in driving style affects the states and behavior of the driver in a highly automated driving situation. Current investigations and this work therefore focus on the driver's state and reaction to a take-over request in a highly automated system.

Tools are needed to continuously and non-intrusively assess the actual workload and vigilance state of the driver, to detect potentially dangerous changes in the demand of driver's attentional resources while engaging in non-driving secondary tasks. Physiological measures could be used to detect workload changes and vigilance level before critical events in highly automated driving. The method of decoding brain signals, i.e. EEG, could be an optimal choice due to its high temporal resolution, sensitivity to different mental states and continuity, compared to other physiological measures.

1.2 Aims of the thesis

As described in Chapter 1.1, variations in driver's mental workload and vigilance level induced by engaging in different non-driving tasks while highly automated driving may have great impact on driver's behavior in take-over situation, and thus influence driving safety. Therefore, the main goal of this work is to continuously and non-intrusively monitor the driver's states including mental workload and vigilance, and to investigate their effect on take-over behavior during highly automated driving. The focus is on the detection of driver's different mental workload niveaus, which are caused by different non-driving tasks during highly automated driving, with the aid of electroencephalography (EEG) data analysis. To fulfill the real-time measuring of workload level in a practical driving scenario, a more simplified workload classifier will be investigated. Meanwhile, if the driver is under high workload or low workload, his/her vigilance level to displayed signals might be different, and thus the detection of vigilance level is also of great importance.

Therefore, in this work the following aspects will be examined:

• Whether it is possible to identify workload levels across different tasks with a simplified EEG-based task-independent classifier;

• Whether it is possible to apply the simplified EEG-based classifier to differentiate workload levels of different non-driving tasks;

• Whether it is possible to detect driver's perception of signals when performing non-driving tasks with EEG analysis;

• What influence do different mental workload levels have on driver's vigilance to signals;

• How does the driver respond to a demanding request to take over of the vehicle, under the influence of different workload levels;

• How does the online warning of high mental workload, which is based on real-time analysis of EEG data, affect driver's reaction to the take-over request.

Overall, by examining the driver's mental states during highly automated driving, this work should help to gain important insights for future systems.

1.3 Structure of the thesis

For a better understanding of automation in motor vehicles, Chapter 2 explains its historical development. A description of the technical functionality and a classification of the levels of automated driving is used to understand highly automated driving in relation to other automation levels. Recent applications by different OEMs and tech companies are listed and the scenario of take-over control is introduced to specify the situation investigated. Chapter 3 gives an overview of changes in driver's mental states related to autonomous driving to emphasize the necessity of monitoring them. In particular, the influence of monitoring mental workload and vigilance is presented. The basics of driver mental workload and vigilance, as well as various methods of their measurement are presented in Chapter 4. To examine the driver state during highly automated driving, the electroencephalograph (EEG) was chosen in this work. In Chapter 5, EEG basics and neural correlates of mental workload and vigilance are described. This chapter concludes

with an overview of studies on workload and vigilance detection based on EEG. Chapter 6 presents EEG analysis techniques, including various preprocessing, feature extraction and classification methods.

Chapter 7 presents a study to validate a task-independent workload classifier with EEG in well-controlled laboratory. Chapter 8 introduces the purpose and a pre-study of the second study, which is aimed at detecting the driver's mental workload and vigilance with EEG during highly automated driving in a driving simulator. Application of the task-independent workload classifier in detecting workload of different non-driving tasks during highly automated driving is presented in Chapter 9. In Chapter 10, the analysis of vigilance detection with EEG and the impact of different workload levels on vigilance are described and discussed. Based on the validation of task-independent workload classifier application, an online study of workload detection is presented in Chapter 11. This chapter also describes the influence of online warning, which is according to the workload classification, on the reaction to take-over request. Finally, chapter 12 summarizes the work and provides an outlook on approaches for further work.

2. Autonomous driving

2.1 Introduction to autonomous driving

Since the invention of the automobile, a driving experience which requires less human intervention or even a certain degree of autonomy was always a goal for the industry. Autonomy in this case could be understood as "*self-determination within a superordinate (moral) law*" (Maurer et al., 2016). Applied on the modern approach of autonomous driving, it could be understood that autonomy is given when the vehicle on its own is constantly making decisions in traffic situations which follows the rules and restrictions programmed by the engineers (Maurer et al., 2016). Well-trained carriage horses were able to achieve a certain autonomy with skills like following a fixed route, avoiding or stopping in case of obstacles, and sometimes even taking over from a groom which is not capable to continue leading the carriage to either return home or make a safe stop at the roadside. Compared to horse carriages that the automobile was replacing, the *autos*, which contains the meaning of "self, independent", achieved the independence from horses but also lost those autonomous decision-making abilities of well-trained carriage horses (Flemisch et al., 2003).

The development of automobile brought a certain regression in autonomy of the transportation tool compared to horse carriages, which resulted in a lesser degree of convenience and safety for the driver. Modern autonomous driving approaches are aimed to regain these functionalities and even more external benefits, such as the positive effects on traffic congestion, more efficient land and energy use and the potential for emission reduction (Anderson et al., 2016). But foremost, the goal of autonomous driving is still to reduce the crash risk of the driver and thereby protecting externals like other drivers, pedestrian, animals or properties. According to the World Health Organization, in 2013 worldwide 1.25 million people died in road traffic related accidents and an additional 20-50 million are injured or disabled, causing over US $ 518 billion in cost or over 1% of the Gross National Product in most countries on the national level (Jacobs et al., 2000). While the main causes of traffic accidents by automobiles at the beginning of the 20th century were due to the lack of traffic rules, reckless driving and technical problems, nowadays human error caused by weariness and distraction of the driver are the main factors (Brown et al., 2014). A study from Cicchino and Jessica (2017) found that automobiles equipped with a forward collision warning (FCW) and a low-speed autonomous emergency braking (AEB), were 56% less likely to be involved in rear-end striking crashes with injuries. Other Studies (e.g. Fildes et al., 2015; Isaksson-Hellman & Lindman, 2016) also had similar results that even low-level automation in driving systems are already providing significant results in crash prevention. Due to the effectiveness of the low-level systems and the high portion of human error in crashes, it is reasonable to assume that technologies that enable farther autonomy for the vehicle, will largely reduce road traffic incidents, therefore saving thousands of lives and reducing economic damages.

2.2 Classification of autonomous driving levels

The goal of autonomous driving is to achieve full autonomous mission through the vehicle. But like all other technologies, autonomous driving has to go through a development progress. To differentiate the technologies moving towards full autonomous driving, classifications of the level of automation were defined. Different levels of automated driving can be defined depending on how strong the system intervenes in the longitudinal and lateral control of the vehicle and whether or not the driver needs to monitor the system (Gasser et al., 2012; Driving, 2014).

Most commonly used is the definition of the Society of Automotive Engineers (SAE), which defines five level of automation for vehicles as follows:

• Level 0, no-automation: Features at this level are limited to warnings and momentary assistance, like anti-lock braking systems, blind spot warning and lane departure warning. The driver is fully in control of the vehicle at all times.

• Level 1, “hands on”, driver assistance: At this level, the systems provide either automated steering while the speed is manually controlled (lane keeping assistance) or automated maintenance of the speed while the steering is manually controlled (adaptive cruise control).

• Level 2, “hands off”, partial automation: At level 2, the systems are able to simultaneously control the speed, the steering and the braking. Though theoretically, the driver can operate the vehicle with hands off, most level 2 systems still require the driver to constantly touch the wheel to assure that the driver is able to take over anytime the system fails to operate properly.

• Level 3, “eyes off”, conditional automation: Starting at level 3, the features can be defined as automated driving features. When certain conditions are met, the driver should be able to turn his attention away from the driving task and the vehicle should be able to handle all driving situations under these certain conditions. Still, the driver should be able to take over within some limited time when it is required.

• Level 4, “mind off”, high Automation: At level 4, the driver’s attention is not required at any time when certain conditions are met. When the conditions are not met, the system is still able to safely abort the driving mission without driver intervention.

• Level 5, “steering wheel optional”, full automation: At level 5, the vehicle should be able to automatically drive everywhere in all conditions, a driver will be redundant.

Apart from the definition by SAE, the German Federal Highway Research Institute (Bundesanstalt für Straßenwesen, in short "BASt") has also defined different levels of autonomous driving (Gasser et al., 2012). Following this definition, the Road Traffic Act in Germany refers to Level 3 automation as "highly automated" rather than "conditionally automated", and Level 4 as "fully automated" rather than "highly automated" (as illustrated in Figure 2.1) (Schubert, 2019).

In this framework, in Level 3, the "Highly automated driving (HAD)", the car will be able to drive autonomously over long distances in certain traffic situations, such as on motorways. The driver, however, must be able to take over control within a few seconds, such as at road construction sites.

Level 4 is considered to be fully autonomous driving, although a human driver can still request control, and the car still has a cockpit. In level 4, the car can handle the majority of driving situations independently. The technology in level 4 is developed to the point that a car can handle highly complex urban driving situations, such as the sudden appearance of construction sites, without any driver intervention.

The level 5 means driverless, where true autonomous driving becomes a reality: Drivers don't need to be fit to drive and don't even need to have a license. The car performs any and all driving tasks, and thus every person in the car becomes a passenger.

Fortunately, the narrative used for the various levels in these two systems is very similar. More detailed information on this can be found in Table 2.1.

SAE Level	SAE Name	SAE Narrative Definition	Execution of Steering/ Accelerating/ Deceleration	Monitoring of Driving Environment	Fallback Performance of Dynamic Driving Task	System capability (driving modes)	BASt Level
Human Driver monitors the driving environment							
0	No Automation	the full-time performance by the human driver of all aspects of the dynamic driving task	Human Driver	Human Driver	Human Driver	N/A	**Driver Only**
1	Driver Assistance	the driving mode-specific execution by a driver assistance system of either steering or acceleration/deceleration	Human Driver and Systems	Human Driver	Human Driver	Some Driving Modes	**Assisted**
2	Partial Automation	part-time or driving mode-dependent execution by one or more driver assistance systems of both steering and acceleration/deceleration. Human driver performs all other aspects of the dynamic driving task.	**System**	Human Driver	Human Driver	Some Driving Modes	**Partially Automated**
Automated driving system ("system") monitors the driving environment							
3	Conditional Automation	driving mode-specific performance by an automated driving system of all aspects of the dynamic driving task – human driver does respond appropriately to a request to intervene	System	**System**	Human Driver	Some Driving Modes	**Highly Automated**
4	High Automation	driving mode-specific performance by an automated driving system of all aspects of the dynamic driving task – human driver does not respond appropriately to a request to intervene	System	System	**System**	Some Driving Modes	**Fully Automated**
5	Full Automation	full-time performance by an automated driving system of all aspects of the dynamic driving task under all roadways and environmental conditions that can be managed by a human driver	System	System	**System**	Some Driving Modes	**Driverless (Autonomous)**

Figure 2.1 Levels of vehicle automation, which are modified based on definitions from SAE (Driving, 2014) and BASt (Gasser et al., 2012)

2.3 Development of autonomous driving

Since the invention of the automobile, making the vehicles safer and more convenient through automation was always a target of the industry. In the beginning of the 20th century, engineers envisioned an automated highway, in which automobiles would operate autonomously through dedicated tracks, magnetics or radio control (Bimbraw, 2015). At the World's Fair in 1939, Norman Bel Geddes presented his vision with cars that communicate with detector circuit buried in the highway through radio to achieve autonomous driving. His vision was farther developed by RCA Labs and GM. From the 1950s to 1960s, RCA and GM collaborated on a series of experiments and demonstrations, and showcased working prototypes with electronical guidance systems that could enable the vehicle to go through the automatic highway driverless. But due to the high reliance of the infrastructure and the consequently high public cost, the demonstration couldn't convince the federal government, thus lost funding and finally faded out. There were other smart road projects developed in several other countries that also failed because of similar reasons.

Obviously, the dream of autonomous car has to be economical, therefore, the new technology should be working without major infrastructure investments. Smart cars were needed instead of smart roads. From the 60s, engineers from around the world began to implement vision-based navigation solutions. Stanford Artificial Intelligence Laboratory developed an electric cart on four bicycle wheels with a camera that uses AI and machine vision to navigate through unfamiliar environment. Although the original purpose of the cart design is for a radio controlled moon cart, the use of onboard sensors and camera was an important development for autonomous driving[1]. In 1977, Japan's Tsukuba Mechanical Engineering Laboratory developed a prototype that achieved a top speed of 30 kilometers per hour driving autonomously using computer vision to analyze the surrounding environment. It is widely regarded as the first stand-alone autonomous vehicle to transport passengers. However, the vehicle still required the support from white markers on the road for guidance[2].

---

[1] Source: https://computerhistory.org/blog/where-to-a-history-of-autonomous-vehicles/

[2] Source: https://computerhistory.org/blog/where-to-a-history-of-autonomous-vehicles/

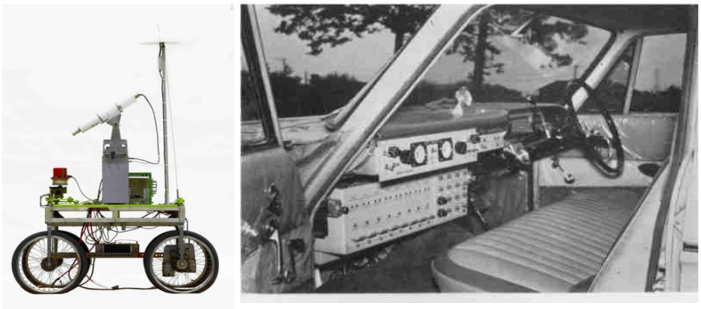


Figure 2.2 An electric cart developed by Stanford Artificial Intelligence Laboratory in 1960s (left, Credit: @ Mark Richards), and a prototype of autonomous driving developed by Japan's Tsukuba Mechanical Engineering Laboratory in 1977 (right[3])

With the fast development of computer technology, from the 1980s to the early 2000s fundamental research in autonomous driving was speeding up around the world. Several landmark projects such as the ALV projects funded by the United States' Defense Advanced Research Projects Agency (DARPA), EUREKA Prometheus Project by Mercedes-Benz and Bundeswehr University Munich, brought significant progresses in the area. In 1986, "VaMoRs", an autonomous Mercedes van equipped with camera and several other sensors designed by the team led by Ernst Dickmanns at Bundeswehr University Munich achieved a top speed of nearly 100 kilometers per hour on roads without traffic (Dickmanns, 2002). During this period, Carnegie Mellon University's NavLab developed 11 versions of the NavLab autonomous vehicles (Thorpe et al., 1991; Pomerleau, 1993). In 1995, NavLab 5 achieved 98.2% semi-autonomous driving on a 5000 km tour across America, where the steering was controlled by a neural-network based system but the throttle and brakes were operated by a human. Similar progress was also made at the University of Parma under the ARGO project, which used two low-cost black and white cameras and stereoscopic vision algorithms on a modified Lancia Thema. The vehicle drove 1900 km on an unmodified highway in northern Italy with an average speed of 90 km/h operating fully automatic for 94% of the distance.

Around the 2010s, autonomous driving became a worldwide technology trend, involving growing interest from research, industry and capital. Google was one of the earlier movers to bet on the autonomous driving future with the goal to provide a driverless passenger fleet. In 2012, Google's experimental self-driving car was the first licensed by the US Department of Motor Vehicles. Till 2020, the self-driving cars by Waymo, which is originated from Google's autonomous driving project, drove over 20 million miles on public roads.

[3] Source: https://computerhistory.org/blog/where-to-a-history-of-autonomous-vehicles/

Although Waymo was given the permission to transport passengers in its vehicle in July 2019, a safety driver is still required.

Tesla was one of the earlier automakers to push their autonomous driving system to the market, but also one of the most controversial. Although Tesla called its system "Fully Automatic Driving", on SAE's scale it only qualifies for level 2 autonomy. This careless misdirection of the customer came with consequences and led to the first fatal accident involving an autonomous car. Several more fatal cases appeared in the following years, evidently caused by drivers that treated the Tesla system as a higher autonomous level system and couldn't take over at emergencies.[4]

Although working on the subject for far longer, traditional OEMs (Original Equipment Manufacturers) are treating the autonomous driving with much greater care. Instead of directly rolling out fully automatic systems, OEMs are gradually adding new sensors and features to their products. In 2018 Audi launched its new Audi A8 equipped with a Lidar enhanced system[5]. Initially, the aim for this model was to reach level 3 on the SAE scale, but due to limitations in the sensors, Audi only allowed it to be used as a traffic jam assistant system on level 2 automation. Until now, there are no true level 3 autonomous cars available for the customer. Various OEMs announced to lunch commercially available level 3 cars in 2021. To provide a reliable level 3 car, take over control is one of the most crucial problems to solve.

---

[4] Source: https://en.wikipedia.org/wiki/List_of_self-driving_car_fatalities

[5] Source: https://www.cnet.com/roadshow/reviews/2019-audi-a8-preview/

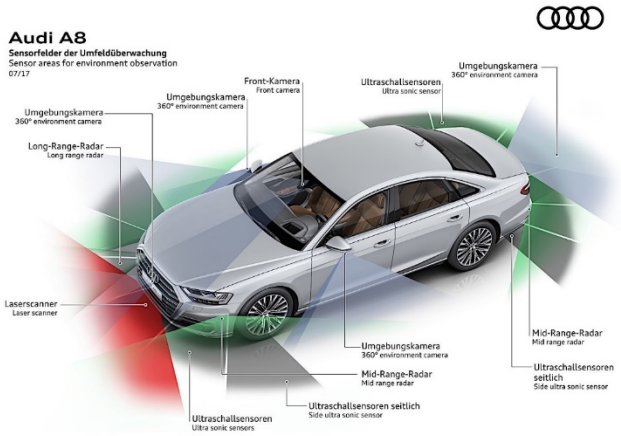


Figure 2.3 An Audi A8 launched in 2018 with Traffic Jam Assistant System[6]

2.4 Take-over control in highly automated driving

In the recent decade level 1 and level 2 automation features such as adaptive cruise control or self-parking became more and more common in commercial automobiles, and millions of driven kilometers are proving their reliability. But in its early development, there were also discussions from the perspective of human factors, if these increased automations could already impair the driver's take-over ability when the system is failing (Merat et al., 2014). Problem in human machine interaction associated with automation was always observed in other areas. When operating automated systems, it could occur that the human operator monitoring the system experiences changes in vigilance and complacency, assumes itself in a passive role in controlling the system, and the quality or form of the feedback from the system could change compared to manual systems (Endsley & Kiris, 1995). These factors can lead to a human out-of-the-loop (OOTL) performance (Kessel & Wickens, 1982), that cause delay in the detection of the problem and delay for the operator to reorient and understand the state in order to take over appropriately, thus affecting the effectiveness of the required take over action. To solve these problems, there has been a human-centered approach on automation (Billings, 1991 & 1997). The automation system, while still performing the tasks autonomously, should have the function to retain the human operator's attention, through tasks that involve the human operator in the control loops. For example, adaptive cruise control and Tesla's autopilot use

[6] Source: https://www.audi-mediacenter.com/en/technology-lexicon-7180/driver-assistance-systems-7184#traffic-jam-assist

features including pressure sensors on the steering wheel and cameras to detect if the driver is still involved in the control loop.

During highly automated driving (level 3 automation), the driver is allowed to be absent from the control loop and has the possibility to shift attention from driving to non-driving tasks while driving automated. This is under the condition that the driver is able to take over control of the system within a reasonable amount of transition time before reaching the system boundary. There are various situations which the system cannot handle and where the driver has to take over. For example, when the system reaches its boundary due to sensor or actuator limitations, or ambiguous environment observations (Merat et al., 2014). One typical example of system limits during highly automated driving are missing or ending lines on motorways, which are normally necessary for estimating the vehicles future path reliably (Damboeck et al., 2012). Besides, false alarms, i.e. take-over requests although the situation is safely managed by the automation, are to be expected. In these cases, a safe transition from highly automated to manual driving has to be ensured and poses one of the major challenges to highly automated driving.

3. Driver states monitoring in HAD

3.1 Importance of driver monitoring in HAD

Today, highly automated driving systems have not been largely designed to replace the driver completely, however, it shifts the driver's role from an operator to a system supervisor. In essence, drivers are required to monitor the driving scene continuously and regain control of the vehicle when it exceeds the limits of the automated system (Merat et al, 2012). In this situation, drivers should not only watch out for other traffic or the surrounding environment, but also take responsibility of the control function of the system.

We might tend to think that automation will eventually relieve human from the task of driving. However, the fact is that even with highly automated systems, the contribution of the human is crucial. The task of monitoring the driving and immediately taking over control of the vehicle when necessary, could even be more demanding than continuous manual driving (cf. Bainbridge, 1983). Although it might not be onerous to monitor the automation and take over control of the vehicle on a quiet and simple road, the interaction with a more complex traffic environment is remarkably different and could be rather tough. For example, when the highly automated vehicle is confronted with other non-automated vehicles or pedestrians performing unexpected actions, and a collision must be avoided, it is especially demanding for the driver to ensure regain of control of the driving task (Merat et al., 2012). If the driver really needs to intervene in emergent situations that cannot be handled by the system, it is a great challenge to ensure a safe take-over of the automated vehicle.

Moreover, during highly automated driving, it's not possible that the drivers just sit and stare, whereas, they might engage in different tasks to keep themselves at a pleasant arousal level (Gold et al, 2015). In the situation of automated driving, drivers tend to be involved in non-driving tasks (Jamson et al., 2013). According to the study of De Winter et al. (2014), there was an increase of 261% in secondary task engagement during highly automated driving, in comparison to manual driving.

Under these circumstances, it is even more challenging to ensure a safe take-over when the driver is engaged in non-driving tasks during automated driving. For this purpose, finding aspects influencing driver's readiness to take over and knowing about the take-over process is of great importance (Zeeb et al, 2015). In order to enhance safety, comfort and efficiency of automated driving, intrinsic human factors challenges must be overcome. Neale & Dingus (1998) stated that "*the hardest problems associated with an Automated Highway System (AHS) … are 'soft'; that is, they are human factors issues of safety, usability, and acceptance, as well as institutional issues. These are problems that are many times more difficult to*

overcome and must be overcome, largely, in parallel with the traditionally 'hard' technological issues" (p. 111).

This concern from a human factors perspective is that the limitation of driver-vehicle-interaction during automated driving may take drivers out-of-the-loop (OOTL), which Endsley and Kiris (1995) argue is a state produced by limited human-system-interaction, leading to the operator's loss of awareness of the system. Highly automated driving induces out-of-the-loop states when the driver is not aware of the driving functions and road situations. Automation alone can take the user out of the loop, and lead to deleterious performance in cases of system failures or boundaries (Brookhuis et al., 2001; Endsley & Kiris, 1995), let alone automated driving with non-driving tasks. For a driver distracted by non-driving tasks, the problem on the out-of-the-loop performance is assumably even more complicated (Zeeb et al, 2015). Being out of the loop will potentially result in overreliance, erratic mental workload, reduced situation awareness, skill degradation, and an inadequate mental model of automation capabilities (Endsley & Kiris, 1995; Parasuraman et al., 2000).

Since the out-of-the-loop state can result in deteriorated performance, it has been suggested that highly automated driving should be designed so that drivers remain engaged and in the loop for optimal performance, and have the ability to regain control of the system when limitations are reached (Louw et al, 2015). Especially on higher automation levels, it is of utmost importance to ensure that drivers are in the loop when required, so that they are able to react properly in a potential critical situation.

To achieve this, Merat and Jamson (2009b) recommended that frequent updates about the drive's state help drivers to better remain in the loop and resume control of the vehicle in a safer manner (Morgen et al, 2016). In order to evaluate drivers' capabilities to take over, it is necessary to monitor their mental states in real time. It has been proven that models assessing and monitoring cognitive states have a significant impact on enhancing the overall performance and productivity (Parasuraman et al., 2005). If the driver's state is detected as for example drowsy, overloaded, or distracted, it is of great importance to bring the driver back into the loop. In such cases, online system feedback is able to facilitate the return into the loop and improve the efficiency in resuming control of the vehicle. Furthermore, it can also increase trust and acceptance towards the automated system (Morgen et al, 2016).

Several models have been used in closed-loop systems or simulation environments to monitor fluctuations in users' mental states, such as workload, vigilance and engagement (Prinzel et al., 2000; Sterman et al., 1993). These systems enable the adaptation based on users' information processing capacity and thus respond specifically to their needs, by means of assessing their internal states (Chaouachi et al., 2011).

3.2 Different mental states

During highly automated driving, a driver is possible to engage in either monitoring the automation driving or performing non-driving tasks. Both situations could have impact on different aspects of the mental state of the driver.

On one hand, in the context of automation, as the drivers' role switches from active control to monitoring the automated system and road situation passively, their ability to recognize crucial factors in the environment or detect system failures, which can be categorized as situational awareness, might be greatly influenced (Llaneras et a., 2013). It is necessary for the drivers to be aware of what is happening on the road, how surrounding traffic is behaving and whether the autonomous vehicle is working effectively. Moreover, a good situational awareness not only concerns the detection of current conditions, but also involves the projection of future actions (Endsley, 1995).

Nevertheless, many studies in aviation imply that automation can affect the situation awareness negatively, which is resulted from overreliance on the system, decreased vigilance and lack of capabilities to understand the system (Merat et al, 2012). Human factors, such as overreliance, skill degradation, as well as the ability to monitor and to detect system failures should be taken into account in the automated driving (Sibi et al, 2016). Overreliance means that the human driver totally relies on the performance of automation and counterchecks the automation status insufficiently. Decreased vigilance means the driver is less ready to detect or respond to certain changes occurring at a random time. Similarly, lack of capabilities means the driver lacks the ability to detect or react at certain time.

The passive monitoring and monotony during highly automated driving can lead to vigilance reduction. Generally, human drivers are bad as monitors for a long time and are likely to be bored if there are very few stimulants (Goncalves et al, 2015). Highly automated driving is a passive situation, in which the drivers only need to monitor the system. It can be thus expected that automation promotes the emerge of drowsiness due to the passive role of drivers as an observer. The switch of driver's role from an active operator to a passive observer and the resulting low demands while driving can induce fatigue and drowsiness, and reduce driver's vigilance level.

In addition, higher automation levels can prevent drivers from receiving feedback within a proper time period, decline the understanding of the process under control, and lead to degraded event detection and response (Wickens, 2008; Saffarian et al, 2012). An automated system might impact drivers' information processing capabilities, and result in deteriorated performance in reacting to the system if necessary.

Another problem of high levels of automation might be the subjectively perceived decrement of workload, causing the driver to interact more with non-driving activities. Their attention will be diverted from monitoring the automation and driving scenario to other non-driving activities in the vehicle. Such distraction might degrade reaction ability and increase reaction time in urgent situations of system failures and when the driver has to take over (Rauch et al, 2009).

From this perspective, we must take the situations of engaging in non-driving tasks into account as well. As before mentioned, an important concern is distraction from the driving scenario. When performing non-driving tasks, the driver might be visually diverted, which means driver's gaze is not fixed to the front roadway, but to other in-vehicle elements on the side such as Infotainment or other driving-irrelevant tasks, such as checking the messages on cellphone. Apart from this, even when the driver seems to be staring at the driving scenario, internal inattention i.e. mind wandering is also likely to happen. Almost all drivers have occasionally drifted their mind towards internal thoughts (Galéra et al., 2012), which could further result in dangerous driving behavior or even traffic accidents (Qu et al., 2015). During highly automated driving, distraction is an inevitable problem inducing poorer responses to critical incidents, and thus a timely detection of drivers' inattention could help to ensure a safe take-over.

Furthermore, mental workload (definition of mental workload see Chapter 4.1.1) is of great interest as it also has impact on drivers' performance during take-over. Overload as well as underload can both impair drivers' performance (Young & Stanton, 1997). This can be observed while drivers engage in a non-driving task, as an inappropriate niveau of workload could cause a decrease in driver's performance in take-over situations (Gold et al, 2015).

3.3 Aspects investigated in this thesis

As summarized before, there are many factors influencing drivers' performance before and during take-over situations during highly automated driving. The work in this thesis intends to detect drivers' mental states noninvasively and in real time. In consideration of existence of previous studies and practicability of certain states' estimation, we decided to concentrate on investigating mental workload and vigilance levels during highly automated driving in this work.

To investigate the influence of automation, many studies emphasize on the change of the mental workload and vigilance of the driver due to the support of the automated system. An automated system, which relieves the human operators but does not replace them, brings both risk and opportunity. It has been frequently

pointed out that automation may lead to changes in workload and simultaneous fluctuations in vigilance (Endsley & Kiris, 1995; Warm et al., 2008).

Highly automated driving and advanced driver assistance systems (ADAS) could increase workload comparing to manual driving, since the driver has to remain vigilant and monitor the automation status. On the other hand, highly automated driving, and to a lesser extent advanced driver assistance systems, could also reduce workload, as the driver is not so engaged in the cognitive activity with regard to manual driving and is relieved from the physical activity of operating the steering wheel and pedals (De Winter et al, 2014).

Automated systems are potentially able to relieve the human driver of tasks that are complex, dangerous, or temporally demanding. Meanwhile, drivers intend to engage in activities other than driving when using automated systems. These driving-unrelated activities require significant cognitive processing resources which may affect the driver's ability to retake control if necessary. Thus, it is of great importance to monitor drivers' mental workload in performing non-driving activities when the vehicle drives autonomously.

To prevent operator errors and to allow system intervention in the way of predicting decline in performance, an accurate measurement of mental workload plays a crucial role. Furthermore, this measurement also promotes a better understanding of driver's mental processes when designing the automated driving system (Sibi et al, 2016).

As the ability of detecting minor changes in the environment of driving is very crucial to safety, vigilance plays an important role in driving. A decrease in vigilance due to either long and monotonous drives or distraction by other non-driving activities, is a main cause of traffic accidents (Thiffault & Bergeron, 2003).

On one hand, being actively engaged in a task induces a depletion of mental resources, and consequently leading to a decrement in vigilance according to the resource theory (Wickens, 2008). Although non-driving tasks could mitigate the drowsiness induced by monotony of automated driving, these tasks might decrease drivers' vigilance level as their attention are distracted. This decline of vigilance could lead to inattentional misperception of take-over request. It is of significant importance to maintain a suitable vigilance level to guarantee sufficient performance before and within take-over situations.

On the other hand, task underload and monotony in automated systems are also reasons for attentional loss and performance decrements. Based on the mindlessness theory (Robertson et al., 1997), which is a theoretical explanation of vigilance decrement, monotony and boredom in such situations can cause task disengagement and mind wandering with task-unrelated thoughts, so that attentional resources are diminished and attentional lapses appear. This could correspondingly lead to deteriorated performance such

as longer reaction times and greater variability of reaction times (Koerber et al., 2015).

Accident data as well as experimental studies have shown that vigilance fluctuations have a significant negative effect on driving safety, especially under monotonous conditions (Dinges, 1995). Therefore, in the future applications, it is of great necessity to realize a reliable, real-time estimation of vigilance state and more specifically, of the driver's reactivity (Schmidt et al., 2007).

4. Mental states investigated in this thesis

4.1 Mental workload

Mental workload is an important part of an operator, especially in safety-critical environments. Mental workload is having a great influence on performance in many tasks, such as in learning, driving, or monitoring (e.g. Berka et al., 2007; Kohlmorgen et al., 2007). In order to ensure the safety, health, comfort, and long-term productive efficiency of the operator, a reasonable goal is to regulate task demands so that they neither underload nor overload an individual. Therefore, a reliable estimation of the actual mental workload during the execution of the task is highly valuable (Wickens, 2008).

4.1.1 Definition

The concept of *mental workload* originated from Industrial and Organizational Psychology in the 1940s (Manzey, 1998), and was introduced to optimize human-machine systems (Bornemann, 1942). Since then, various definitions have appeared, whereas there is neither a common definition of mental workload nor a specific measurement.

According to De Waard (1996), "*a simplistic definition of workload is that it is a demand placed upon humans*", which attributes workload solely to an external source. However, in O'Donnell & Eggemeier (1986)'s definition, workload is the "*portion of the operator's limited capacity that is actually required to perform a particular task*". Similarly, Eggemeier et al. (1991) defined mental workload as "*portion of operator information processing capacity or resources that is actually required to meet system demands*". In these definitions, workload is not only task-centered, but it also depends on the amount of resources the operator could or would allocate in relation to the demands, which means it is person-specific as well (Meijman & O'Hanlon, 1984; Rouse et al., 1993). A basis for this would be Kahneman's (1973) theory that cognitive resources for human information processing are limited. Therefore, the amount of task demands placed on operator's limited resources is equivalent to mental workload. In Wickens' (2002) assumption mental workload was similarly described as the "*relation between the (quantitative) demands for resources imposed by a task and the ability to supply those resources by the operator*".

There are numerous definitions regarding mental workload. Although there is no agreement upon the definition, different perspectives share some common understandings of the concept that mental workload is the ratio between task demands and a person's limited mental resources (Ruff, 2017). When task demands exceed the capacity, the operator will be overloaded and performance will decline.

Overload occurs if the demands of a task are beyond the limited attentional capacity of the operator, and results from a combination of task characteristics, such as time pressure and task difficulty. Contrarily, underload occurs when less effort is made to cope with the demands, and is always associated with passivity (Young & Stanton, 2002). Overload and underload can both lead to psychological strain due to a mismatch between demands and capabilities (Byrne & Parasuraman, 1996).

In the context of driving, mental workload plays an important role in the driving safety. Engaging in several tasks simultaneously while driving could lead to higher mental workload and distraction as well (De Waard, 1996). Many factors, including individual factors, such as driving experience and age, and environment factors, such as road conditions, may affect driver workload and driving performance. With the increasing implementation of in-vehicle technologies (IVTs), such as the navigation system, human factor issues concerning these devices should also be put more emphasis on. It has been pointed out that the use of IVTs may result in driver overload, especially in a complex driving situation (Verwey, 1990). Therefore, an accurate assessment of mental workload is crucial and helps maintain the driving safety.

4.1.2 Measurements

There are diverse tools for the evaluation and prediction of mental workload. Most of the workload assessment techniques can be divided into the three main categories: performance measures, subjective measures, and physiological measures (O'Donnell & Eggemeier, 1986; Wierwille & Eggemeier, 1993; Cain, 2007).

4.1.2.1 Performance measures

Performance measures are grounded on the assumption that an increase in task difficulty will raise demands and correspondingly lead to performance decrement. When demands exceed the operator's capacity, the performance degrades from baseline or ideal levels. Thus, performance measures are often applied as important index of workload. Generally, there are two kinds of performance measures: primary task measures and secondary task measures.

Primary task measures

Primary task measures aim to assess directly the operator's performance on the task at hand. In laboratory tasks, we frequently use motor or tracking performance, number of errors, accuracy, speed performance, or reaction time measures as primary task performance measures (De Waard, 1996). These measures allow for a continuous assessment without interference. However, primary task performance is, by its nature, very

task-specific. There is no prevalent primary task measure and it's hard to directly compare between two different primary tasks. Even the same performance decreases could be induced by structurally different task demands (Vidulich & Tsang, 2012). Therefore, it is necessary to combine primary task measures with other workload measures to draw a solid conclusion.

Secondary task measures

The Subsidiary Task Paradigm is often used in mental workload assessments, in which a simultaneous secondary task is added to the primary task. Participants are instructed to maintain the primary task performance. Consequently, secondary task performance varies with difficulty and reflects 'spare capacity' available to the operator, under the assumption (Brown & Poulton, 1961) that the amounts of capacity to perform all tasks are totally undifferentiated. Spare capacity is described here as the unused capacity which is available for the secondary task in the case of unaffected single task performance (De Waard, 1996).

Secondary task measures are often used for driver workload detection, especially for the workload assessment when the in-vehicle secondary task interferes with the primary driving task (e.g. Martens & van Winsum, 2000; Mattes, 2003). However, during highly automated driving, the primary task of driving is no more dominant, drivers tend to pay more attention to non-driving tasks. In such situations, these non-driving tasks turn to be more like primary tasks, and the traditional secondary task measures are not suitable any more.

4.1.2.2 Subjective measures

Subjective measures assume that an increased power expense is related to perceived effort and can be appropriately assessed by individuals (Rubio et al., 2004), and they attempt to quantify the individual interpretations and judgements (Cain, 2007). Subjective measure has been applied in many research areas such as assessment of workload, effort, and emotion, and is considered to perform best because of its subjectivity. While operating any task, no performance degradation doesn't mean the invariability of effort used, whereas, the operator could be aware of an increase of effort used for task completion, even task performance remains stable (Muckler & Seven, 1992). In this sense, subjective measure plays a vital role in assessing mental workload.

There are plenty subjective rating scales and many of them have been frequently used, such as the NASA Task Load Index (NASA-TLX, Hart & Staveland, 1988), the Subjective Workload Assessment Technique (SWAT, Reid et al., 1981), the Driving Activity Load Index (DALI, Pauzie, 2008) especially for automotive context, and the SEA-Scale (Eilers et al., 1998). In this thesis, two questionnaires – the NASA-TLX and the

SEA-Scale – were applied to assess mental workload.

NASA Task Load Index (NASA-TLX)

The NASA-TLX is very often used for subjective measurement. It assumes that the workload is influenced by several factors, including mental demand, physical demand, temporal demand, performance, frustration level and effort, and the combination of these dimensions is able to represent the experienced "workload" during most task implementation by most people (Hart & Staveland, 1988). Initially, the NASA-TLX was designed for use in aviation, and nowadays this tool is also widely used in the context of driving and related designs, like In-Vehicle Information Systems and Advanced Driver Assistance Systems.

The traditional NASA-TLX consists of the process of 15 pairwise comparisons between two of the six dimensions, before calculating the overall workload rating. This process is rather time-consuming, and in order to simplify the rating procedure, Byers et al. (1989) proposed a Raw Task Load Index (Raw TLX) which does not require pairwise comparisons but only computes a simple average of the six NASA-TLX scales. Byers and his colleagues proved that the traditional and new version of NASA-TLX had comparable means and standard deviations, and the correlation is high ($r = 0.95$), thus they recommend using the Raw TLX as a much simpler alternative. Hence in this thesis, due to time limitations, we will use the Raw TLX way to compute the overall workload ratings and all the "NASA-TLX" mentioned below refer to the Raw TLX version.

SEA-Scale

SEA-Scale (in German: Skala zur Erfassung von subjektiv erlebter Anstrengung) was developed by Eilers et al. (1998) and its original version is in German. SEA-Scale was designed for measuring the subjectively perceived workload and it is one of the measuring instruments for determining the mental efficiency of a product and its use in terms of usability [ISO 9241- 11]. On the basis of studies in other languages for the subjective measure of mental effort, a point value scale with values from 0 to 220 was established, which has markers with certain terms at certain point values (see Appendix A). With SEA-Scale, the overall mental workload is assessed on a metric scale. SEA-Scale is widely used in workload measurements in German context and is especially appropriate for efficiency evaluation of Advanced Driver Assistance Systems (Pataki et al., 2014).

4.1.2.3 Physiological measures

Although the most frequently used method for mental workload assessment are subjective measures due to

their ease of application and low costs, these methods are typically intrusive and prone to social expectance and are hence unsuitable for continuous workload assessment. Alternatively, psychophysiological measures can be used to measure correlates of mental workload non-intrusively and continuously. These measures provide maximal objectivity since people can barely control the outcome of the recordings (Hagemann, 2008).

Physiological measures assume that the mental workload can be measured by means of the level of physiological activation. A variety of physiological parameters are reported to be differently sensitive to changes in the task load (De Waard, 1996). Generally, physiological indicators fall into two categories associating with two anatomical distinct structures: central nervous system (CNS) and peripheral nervous system (PNS). CNS includes the brain, brain stem and spinal cord cells, while PNS can be divided into the somatic nervous system and autonomic nervous system (ANS). The somatic nervous system relates to the activation of voluntary muscles, and ANS controls internal organs unconsciously. ANS measures include pupil diameter, heart rate and respiratory, electrodermal activity (EDA) and hormone level measures. CNS measures consist of electroencephalogram (EEG), magnetic and metabolic activity measures of the brain and eye movements. Apart from these two categories, there's another group of measures on peripheral responses, which includes spontaneous muscle activity and eye movements (O'Donnell & Eggemeier, 1986).

Physiological measures can be used in various operational environments, such as car driving. In the following part, several frequently used physiological measures for workload assessment in the context of driving will be introduced.

Cardiac measures

Heart rate (HR) and heart rate variability (HRV) are commonly used cardiac indicators from the electrocardiogram (ECG), which measures electrical activity caused by depolarization and polarization of the heart muscle and reflects the electrical impulses induced by heart contraction. HR and HRV are widely used for the representation of mental workload, since they are easy to record and less sensitive to artefacts (Kramer, 1991). In general, an increment of mental workload leads to HR increase and HRV decreases. However, some researchers (e.g. Lee & Park, 1990; Jorna, 1992) demonstrated that physical load could also cause the changes in HR and HRV, and other factors, such as emotion and alcohol, could also influence these indices.

Ocular measures

Eye activity can be measured unobtrusively with various equipment, such as eye-tracker, video camera, and

electrooculogram (EOG). Fixation is one of the most commonly used indicators of eye activity in mental workload assessment. Rötting (2001, p.68) defines fixation as a "state, in which the eye is in relative stasis to its object of interest". The most outstanding fixation characteristic is fixation duration. Normally, an increase in workload is accompanied by increased fixation duration. Whereas, when the information needed to perform a task is visually complex, fixation durations could decrease (Holmqvist et al., 2011). Another prominent ocular measurement is pupil diameter, which has long been researched to indicate the change in cognitive load (Hess & Polt, 1964; Beatty, 1982). Pupil diameter increases with increment in perceptual, cognitive and response-related processing demands, and its variation can be observed regardless of task modalities. However, Wilson and Russell (2004) pointed out that under overload pupil diameter can become unresponsive to changes or even reverse its response. Besides, pupil diameter is also sensitive to light changes (Wyatt & Musselman, 1981).

In general, although ocular measures are sensitive to mental workload, they are also highly influenced by other factors, particularly fatigue (Cain, 2007).

Electroencephalogram (EEG)

An electroencephalogram (EEG) is a method recording the electrical signals of the brain, and is typically noninvasive, with electrodes placed along the scalp. With its advantages in speed and sensitivity, EEG technology has been widely used in psychophysiological and human factor research (Kramer, 1991). For workload assessment, EEG is particularly useful because of its high temporal resolution, portability, and continuity. According to former studies (e.g. Klimesch, 1999; Lei, 2011), brain activity reflects an increase of mental workload by an increase in frontal theta power and a decrease in parietal alpha power. Moreover, Hogervorst et al. (2014) have demonstrated that in comparison to other psychophysiological measures, such as eye related measures and ECG, EEG outperforms them in online classification accuracy. EEG is also suggested to be the most sensitive or promising indicator of workload due to its high temporal resolution in reflecting subtle shifts in workload (Berka et al., 2007). Therefore, EEG was applied in our experiments to index mental workload. A more detailed introduction to EEG technology will be given in Chapter 5.

4.2 Vigilance

4.2.1 Definition

According to Mackworth's (1957) definition, vigilance describes "*a state of readiness to detect and respond to certain specified small changes occurring at random time intervals in the environment*". Parasuraman

and his colleagues (1998) also gave a broader definition of vigilance as "*the ability to sustain attention to a task for a period of time*", which may refer to a general state of wakefulness that is characterized as arousal or alertness.

Vigilance is directed toward specific goals, such as monitoring or detecting anomalies, and means that the observer needs to remain alert to potential critical signal occurrences for prolonged periods. Vigilance is a crucial factor in safety and can be affected by many aspects. Several studies suggested that vigilance problems might be related to the increasing degree of automation in technological systems. For example, in aviation the use of highly automated flight decks may lead to cognitive underload in pilot performance accompanied with boredom and mind wandering, which may cause poor detection of vital stimuli (Pattyn et al., 2008). Individual factors also play a role in vigilance-related decrements. Prinzel and his colleagues (2001) demonstrated that in an automation environment boredom-prone people were less able to detect malfunctions than those who were not prone to boredom.

Thiffault and Bergeron (2003) divide the factors influencing vigilance into exogenous and endogenous aspects, depending on whether they originate from within the organism or they are caused by characteristics of the task performed.

In the context of driving, sleepiness is influenced by circadian and homeostatic variables and is thus an endogenous factor, which can be mitigated with sleep. Monotony of the road environment or in driving itself such as in automated driving may influence vigilance as an exogenous factor. Monotony in driving is not only caused by a lack of alerting stimulation but also by a high predictability of the situation. Independent of whether monotony, sleepiness or any other factors reduce drivers' vigilance, fluctuations in vigilance in general and a vigilance decrease in particular could induce a serious risk to safety. Hence, it is of great importance for the drivers to be able to judge their vigilance state correctly for the sake of traffic safety (Schmidt et al., 2009).

From the definition of vigilance, the ability to detect stimuli or changes is a key factor in remaining vigilant or alert. To assess vigilance level, reaction performance such as reaction time or detection rate is widely used. In this work, the investigated vigilance level mainly refers to driver's reaction readiness to detect and respond to the take-over request stimuli.

4.2.2 Measurements

Performance measures

Generally, a decrement in vigilance is usually defined by a performance reduction. Based on this, performance measures have high face validity for assessing vigilance states. Performance measures for vigilance evaluation include the performance measure of task itself and measures of secondary task performance. In the context of driving situations, it's suggested to avoid degrading driving performance too seriously in order to maintain the safety in driving. Thus, it would be harder to directly vary the driving task conditions when conducting experiments. In contrast, implementing and modulating secondary tasks are more viable and more likely not to interfere with the driving task. Several studies have demonstrated that a secondary task is sensitive to even minor changes in vigilance and predictive of variations in ability to respond to unforeseen events (e.g. Laurell & Lisper, 1978; Graw et al., 2004; Schmidt et al., 2009).

Subjective measures

There's a large debate on the reliability of assessing vigilance with subjective measures. Although several studies suggested that people are generally able to perceive the loss of performance and deteriorated vigilance in tasks such as in driving, especially under sleep deprivation and at night (Baranski, 2007; Horne & Baulk, 2004). However, a series of studies have shown that subjective ratings are not sufficient to serve as an accurate, reliable and valid indicator of vigilance, in the aspects such as reaction time, driving performance, or sleep propensity (Belz et al., 2004; Philip et al., 2003). Self-assessment is rather poor in measuring the change in performance and exact extent of vigilance variation.

Physiological measures

Electroencephalogram (EEG)

A commonly used physiological method to measure vigilance level is electroencephalography (EEG). To investigate the physiological markers of high- and low-vigilance states, several studies have examined the association of certain features of EEG data with operator performance.

The most often assessed feature is the amplitude of the P3 component of event related potential, which is induced by stimulus and has been shown to be sensitive to changes in vigilance (Koelega et al., 1992). The P3 amplitude can be interpreted as a measure of the processing depth of stimuli. Apart from ERP, rate of the spontaneous alpha spindle, which is a feature derived from the alpha-band (8–12 Hz), has also been proven to correlate with variations in vigilance (Papadelis et al., 2007).

Ocular measures

Another vigilance measurement focuses on eye tracking. It has been demonstrated that an increase in blink

frequency and blink duration is related to decrement of vigilance (McIntire et al., 2014). The percentage of eye closure (PERCLOS) for a given time accompanied by a decrease in pupil diameter is also shown to associate with an increment in drowsiness and a decrement in vigilance (Abe et al., 2011). Furthermore, an increase in eye closure occurs in driving when the automation is activated (Buld et al., 2002).

Other physiological measures

In addition, there are some other physiological measures for vigilance assessment, which are not so frequently used. For example, heart rate can also serve as an indicator of the physical activation level and has shown to be sensitive to vigilance changes (O'Hanlon and Kelly, 1977). The functional magnetic resonance imaging (fMRI) approach has also be applied in investigating the actual state of vigilance during task performance (Czisch et al., 2012).

5. Electroencephalography

5.1 Definition

Literally, electroencephalography (EEG) is the graphical depiction ('graphy') of electrical activity ('electro') of the brain ('encephalon'). In other words, it's a medical imaging technique that reads scalp electrical activity generated by the brain.

EEG's history can be traced back to over 100 years ago. An English physician Richard Caton observed the existence of electrical currents in the brain from the exposed brains of rabbits and monkeys in 1875 (Teplan, 2002). In 1924, a German neurologist Hans Berger successfully amplified the brain's electrical activity measured from the scalp with ordinary radio equipment. He announced that weak electrical currents produced in the brain can be recorded without opening the skull, and depicted graphically on a strip of paper. Berger laid the foundations for the applications of electroencephalography afterwards. He was also the first person using the word "electroencephalogram" to describe human brain's electrical activities. Furthermore, he suggested that the brain activity changed according to the functional status of the brain, such as in sleep, anesthesia, and lack of oxygen, and affected by certain pathologic conditions, such as epilepsy (Berger, 1929).

Later in 1934, Adrian and Matthews published the paper confirming the concept of "human brain waves", and more importantly, they identified certain rhythms in the EEG, e.g. regular oscillations around 10 to 12 Hz recorded from the occipital lobes of the cerebral cortex, which they termed "alpha rhythm" (Bronzino, 1995).

Nowadays, EEG is defined as a recording of electrical activities from the scalp surface which are picked up by means of metal electrodes and conductive media (Niedermeyer & Lopes da Silva, 1993). EEG recording is a completely non-invasive procedure that can be iteratively applied to patients, normal adults, and children with virtually no risk or limitation (Teplan, 2002). Because of its relatively simple setup, low cost, speed and sensitivity, EEG technology has been widely used in medicine, neuro- and psycho-physiological measures, and human factor research (Kramer, 1991).

In order to have a better knowledge of EEG, an introduction to how EEG is generated will be presented at first. The electrical activity within the brain arises because nerve cells process and exchange information in the form of electrical impulses. The brain consists of estimated over 10^{11} neurons that are interconnected to form a complex network. Each nerve cell consists of a cell body (soma), a single axon, and various dendrites (see Figure 5.1). Most neurons receive signals via the dendrites and the axon transmits the nerve impulses

from the soma to the downstream neuron. The surface of dendrites consists of innumerable contact points called synapses. The respective axons and dendrites of different nerve cells are connected to each other via these synapses and thus receive or transmit nerve impulses from neighboring cells.

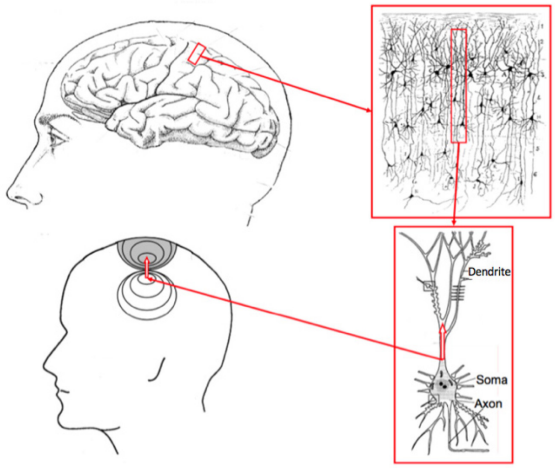


Figure 5.1 Illustration of human brain and the structure of the origin of electrical activity within the brain (source: Welke (2012), which modified based on Hämäläinen et al. (1993) and Vigário et al. (2000))

The electrical impulses that are exchanged between the nerve cells via the axons are called action potentials (AP). These potentials are generated from changes in the resting membrane potential in the cell interior of the neuron. The resting membrane potential is regulated by a protein anchored in the cell membrane, which is also referred to as sodium-potassium pump (Na+/K+-ATPase). Normally, the potential outside of the cell is about 70mV higher than the potential inside. For a neuron at rest, there is a high concentration of Na+ ions outside of the cell while a high concentration of K+ ions inside of the cell. Initially, Na+ channels open and some Na+ ions moves into the cell, causing depolarization. When the depolarization reaches its threshold, more Na+ ions will flood into the cell, and this will create a negative potential outside the cell and a positive potential inside the cell. At this time, Na+ channels close, K+ channels open and repolarization occurs. This leads to an outflow of the K+ ions from inside of the cell, and the cell interior is slowly negatively charged (this is called hyperpolarization), thus recovering to the resting membrane potential. Due to the concentration increase of Na+ ions inside and K+ ions outside of the cell, the sodium-potassium pump will be activated and moves sodium ions (Na+) out of the cell and potassium ions (K+) into the cell. In this process, a neuron is able to open or close special ion channels very quickly.

The emergence of action potential is consequently the result of a complex interaction of the external excitation and the specific properties of a neuron itself. A series of action potentials enables electrical

information transfer between the individual neurons in the brain network. The propagation of action potential among the neurons is based on the connection between the synapses and axons. In order for the action potential to be passed on to the next neuron, the non-conductive synaptic gap between an axon and the dendrite of a downstream nerve cell must be overcome. The electrical action potential is converted into a chemical signal for transmission. The transmitter diffuses to the cell wall of the postsynaptic cell, leading to local depolarization and the generation of an electrical potential called postsynaptic potential. Due to the nature of the synapses, the neuronal circuits differ in excitatory (increasing the action) and inhibitory (inhibiting the action) connections. These connections determine whether and how an action potential is forwarded (Welke, 2012).

Neurons receive and send information based on the change in their chemical and electrical properties. However, only the synchronous activity of a great number of neurons produces measurable electrical activity on the head surface. The electrical activity of neurons in the brain generates currents that reach the surface of the scalp, and EEG provides a non-invasive method of recording the voltage differences of these scalp potentials. Normally, an EEG recording system consists of four parts, the electrodes with conductive media, amplifiers with filters, A/D converter, and recording system. Electrodes read the brain signal from the scalp, and the signal is transmitted from the scalp electrodes to the amplifiers in order to amplify the microvolt signals, which are severely attenuated through the skull, into a range which can be digitalized in an accurate way. The transmitted signals are then changed from analog to digital form by the converter. At last, the recorder device stores and displays the collected data (Teplan, 2002). Therefore, the EEG signal can be continuously sampled at a high rate (normally between 10 to 1000 Hz) to provide a high temporal resolution.

To record the EEG, electrodes are fixed to the scalp with a cap or a mesh, according to the 10-20 system, which was proposed by the International Federation in Electroencephalography and Clinical Neurophysiology (Jasper, 1958) for the arrangement of the electrodes. The electrode positions are distributed in 10% or 20% steps based on characteristic points of the human skull on the scalp (see Figure 5.2). In order to make more precise cortical representation of cognitive processes, these positions are slightly changed (Welke, 2012). The electrodes are thus named in accordance with adjacent brain areas: F (frontal), C (central), P (posterior), O (occipital), and T (temporal), with odd numbers on the left side and even numbers on the right side. In practical use, the recording systems are typically with 32, 64, 128 or 256 electrodes.

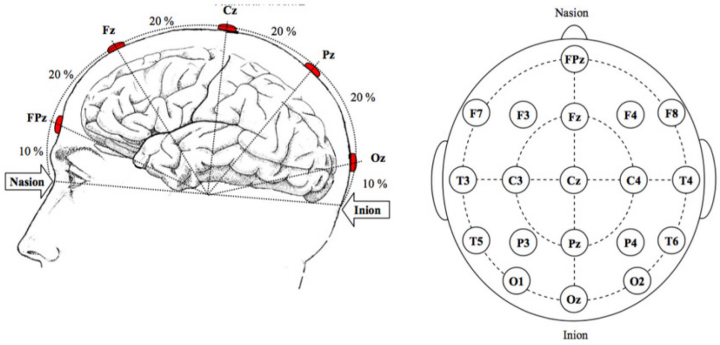


Figure 5.2 Placement of EEG electrodes according to the 10-20 system (left) and electrode positions on the scalp in a top view (right) (source: Welke (2012), according to Jasper's (1958) 10-20 system)

The recorded EEG signal reveals the functional dynamic within the brain, and this dynamic is reflected in changes in the amplitude, frequency (Pfurtscheller & Lopes da Silva, 1999) and phase (Makeig et al., 2002) of the waveforms of the signal. Based on the this, the oscillatory activity of EEG can be categorized into five basic groups: delta (0-4 Hz), theta (4-8 Hz), alpha (8-12 Hz), beta (12-30 Hz) and gamma (>30 Hz) (Bronzino, 1995), as shown in Figure 5.3. Individual's brain state may make certain frequencies more dominant, therefore, the analysis of power spectrum is able to reveal different cognitive states.

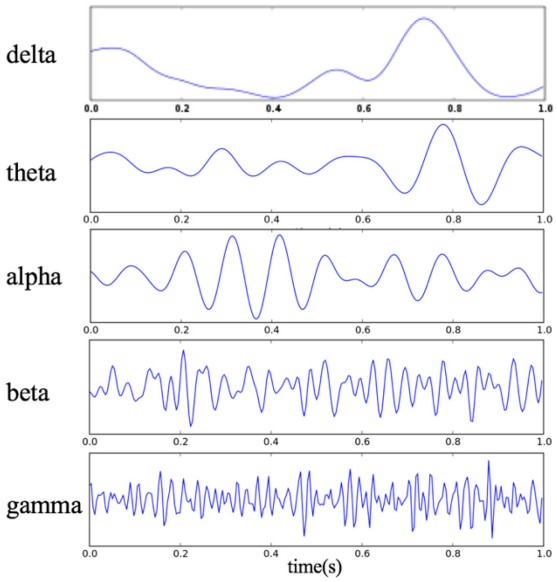


Figure 5.3 EEG frequency components delta, theta, alpha, beta and gamma[7]

Delta is normally seen in the slow-wave sleep of adults and in babies. Delta waves usually represent the onset of deep sleep phases in healthy adults (Rechtschaffen & Kales, 1968). It is most prominent in the frontal region in adults.

Theta is usually associated with drowsiness, arousal, deep relaxation and meditation, and can be seen at the transition stage between wake and sleep (Hagemann, 2008). Furthermore, theta oscillations are also related to the attentional control mechanism in the anterior cingulated cortex (Smith et al., 2001) and often increase with an increment of cognitive task demand (e.g. Gevins et al., 1997).

Alpha can be observed better in the posterior and occipital regions and is usually induced by closing the eyes and by relaxation, and abolished by eyes opening or alerting. Most people are extraordinarily sensitive to eye closing, that is to say, the EEG pattern significantly changes from beta into alpha when closing eyes (Teplan, 2002). High alpha power has been assumed to reflect a state of relaxation or cortical idling (eyes closed but awake).

Beta is often linked to motor behavior and decreased during active movements, and is most dominant in the

[7] Source: https://en.wikipedia.org/wiki/Electroencephalography

frontal area. Low-amplitude beta is closely associated with active, busy or anxious thinking and active concentration (Pfurtscheller & Lopes da Silva, 1999). Moreover, it has been shown to represent visual concentration and attention orientation (Birbaumer & Schmidt, 1996).

Gamma is infrequent during awake states of consciousness (Dooley, 2009). It is associated with many cognitive functions such as attention, learning, memory and language perception (Eulitz et al., 1996; Jensen et al., 2007). Large gamma oscillations are also affected by attentional modulation (Tallon-Baudry et al., 2005).

Apart from the power spectrum analysis, another commonly used approach for the EEG analysis is the event-related potential (ERP). Event-related potentials (ERPs) are the voltage fluctuations that are related in time to some physical or mental occurrence (Picton et al., 2000). ERPs are measured brain responses to a stimulus, which consists of relatively fast positive and negative waves and last for about half a second (Kotchoubey, 2006).

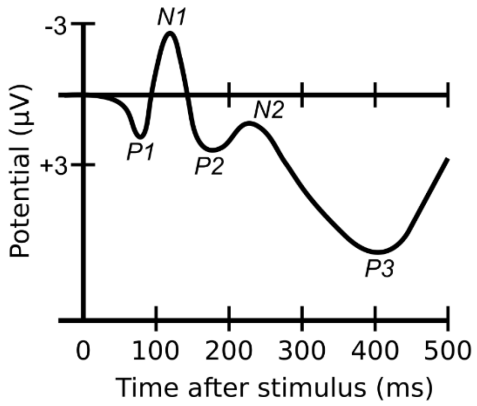


Figure 5.4 Typical Event-Related Potential waveform[8]

As shown in Figure 5.4, the ERP waveforms contain a series of positive and negative voltage deflections and these ERP components are referred to by a letter (P/N) indicating polarity (positive/negative), followed by a number indicating either the latency in milliseconds or the component's ordinal position in the waveform. For instance, a negative-going peak that is the first substantial peak in the waveform and often occurs about 100 milliseconds after a stimulus is presented is often called the N100 (indicating its latency

[8] Source: https://en.wikipedia.org/wiki/Event-related_potential#:~:text=An%20event%2Drelated%20potential%20(ERP,means%20of%20evaluating%20brain%20functioning.

is 100 ms after the stimulus and that it is negative) or N1 (indicating that it is the first peak and is negative); it is often followed by a positive peak, usually called the P200 or P2. The stated latencies for ERP components are often quite variable, particularly so for the later components that are related to the cognitive processing of an event or a stimulus. For example, the P300 component may exhibit a peak anywhere between 250 ms and 700 ms, not always 300 ms after the stimulus. ERP components are supposed to allow for the possibility to obtain information about how the intact human brain processes signals and prepares actions.

5.2 Measuring mental workload with EEG

Detection of mental workload with EEG has been studied by many researchers, and a large amount of studies have validated that EEG changes accompany increases in cognitive workload and the allocation of mental effort.

Most studies gain information from the frequency domain of the EEG in respect of mental workload and have identified the correlation between workload and spectral power in certain frequency bands of EEG, in particular the alpha (8-12 Hz) and theta (4-8 Hz) bands. The spectral components of EEG show typical fluctuations corresponding to changes in mental workload (Parasuraman & Rizzo, 2007). A number of studies have recognized that differences in the alpha and theta band are the most sensitive indicators in EEG when distinguishing between different workload levels. Alpha power is a vital index of mental effort because of its association with resource allocation and workload (Käthner et al., 2014). A decrease of alpha power is often related to increases in task difficulty and mental workload, and this alpha decrease is most profound at parietal regions (e.g. Klimesch et al., 2000; Keil et al., 2006). There is also sufficient evidence for the relationship between theta and working memory load or mental effort. Theta increases as task requirements increase (e.g. Miyata et al., 1990; Klimesch, 1999; Raghavachari et al., 2001), and this theta increase can be observed over frontal electrode positions (Gevins et al., 1998).

Many studies have also reported that both alpha and theta are associated with mental workload (e.g. Brookings *et al* 1996, Gevins *et al* 1998, Fournier *et al* 1999, Gundel and Wilson 1992). An increase of task difficulty and mental workload causes the decrease of alpha and increase of theta power (Klimesch, 1999; Scerbo, Freeman, & Milkulka, 2003). Gevins and his colleagues (1990, 1997, 1998) used an N-back task to vary working memory load and found that frontal midline theta power increases during high task load conditions and attenuated alpha power proportional to the increase of workload (Gevins et al., 1990; Gevins et al., 1997; Gevins et al., 1998). More specifically, we could observe this effect in Figure 5.5, which

presents the spectral power in the 4–14 Hz range at a frontal midline (Fz) and a parietal midline (Pz) scalp location computed from the continuous EEG under conditions of low-load (0-back) and high-load (2-back) of a spatial *n*-back task (data from Gevins & Smith, 2000).

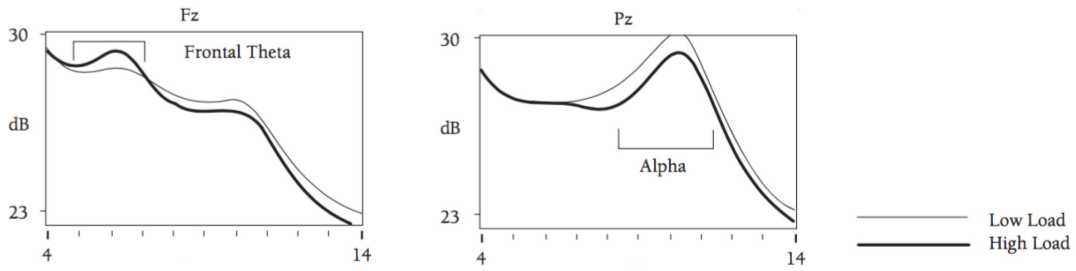


Figure 5.5 Typical changes in spectral power of EEG in the 4-14 Hz range at frontal (Fz) and parietal (Pz) midline electrodes under high and low workload conditions (Data from Gevins & Smith, 2000)

In addition to theta and alpha bands, power in some other frequency bands has also been reported to respond to changes in workload. For instance, the slow-wave activity in delta (<3 Hz) band and high-frequency activity in beta (15–30 Hz) and gamma (30–50 Hz) bands are sensitive to workload variations (McCallum, Cooper, & Pocock, 1988; Sheer, 1989).

Besides EEG spectral components, ERPs potentially convey relevant information on user's workload. Based on the observation that P300 amplitude was reduced in dual-task studies, it was proposed that the P300 amplitude reflects the allocation of processing capacity between concurrent tasks (see Kramer, 1991; Kok, 2001 for reviews). More recently, studies also demonstrated attenuated P300 amplitude under high workload conditions of cognitive tasks or more realistic tasks such as gauge monitoring (e.g. Ullsperger et al., 2001). Apart from the P300, earlier ERP components like the N100, N200 and P1 have been found to associate with task difficulty or workload (Ullsperger et al., 2001; Kramer et al., 1995; Pratt et al 2011), and the late positive or negative slow waves have relations with high workload and amount of resource distribution (see Ruchkin et al., 1990; Rösler et al., 1997; Brouwer et al., 2012 for reviews).

Despite many applications of workload estimation in standardized cognitive tasks in the laboratory, it is also possible to extend these EEG-based methods for workload monitoring in more naturalistic task and situations. Due to its high temporal resolution, simplicity and portability of use, EEG proves to be a suitable technology for use in unrestrained subjects in a wide range of environments, even in real-world work contexts (Parasuraman & Rizzo, 2007). For instance, Smith and his colleagues (2001) investigated the application of EEG in measuring workload of the Multi-Attribute Task Battery (MATB), a computer-based multitasking environment simulating the activities that a pilot might be required to perform. Analogous

methods were used in a study by Gevins & Smith (2003) involving more naturalistic computer-based tasks, such as word processing, taking a computer-based aptitude test, and searching for information on the website. In another study by Pellouchoud and his colleagues (1999), they recorded EEG while playing a videogame or watching others playing, and found higher frontal midline theta rhythm and attenuated posterior alpha band signal in playing the videogame than passive watching or resting conditions.

EEG indices play a significant role in workload detection in realistic simulation and operational environments, whereas, it is even more important to realize the online detection of workload. Although a large amount of studies has demonstrated the feasibility of both spectral power of certain frequency bands and ERPs in workload estimation, the spectral components could convey relevant information in a more straightforward way than ERPs and this information in spectral power can be extracted from the ongoing EEG in real time. For example, Kohlmorgen and his colleagues (2007) applied classifier which computed the power of selected frequency bands to identify high mental workload in a real-life driving situation. They measured workload while driving in real time, and implemented an online adaptation system based on real-time workload estimation, which was proven to be effective in improving driver's performance. In Lei and his colleagues' study (2017), it was able to detect driver's mental workload by EEG in real time and use the result to adapt a secondary task allocated to driver.

5.3 Measuring vigilance with EEG

The most often used feature for vigilance measurement is the amplitude of the P3 component of event related potential (ERP), which is induced by stimulus and has been shown to be sensitive to changes in vigilance (Koelega et al., 1992). The P3 amplitude can be interpreted as a measure of the processing depth of stimuli.

To assess performance- and event-related physiological measures, a classical oddball paradigm has been often used. Amplitudes of the P3 components of ERPs were proven to be sensitive to changes in vigilance state (Schmidt et al., 2007). If a typical ERP sequence is detected, the participant should have responded adequately to the target stimulus. If a pending response is not accompanied by an ERP, the participant might have missed to detect the target stimulus.

Apart from ERP, Bonnefond and his colleagues (2010) observed increases in alpha and theta power during states of low vigilance. An index involving beta, alpha and theta power was also capable of revealing decrement in vigilance in a visual oddball task (Freeman et al., 2004). In addition, rate of the spontaneous alpha spindle, which is a feature derived from the alpha-band (8–12 Hz), has also been demonstrated to associate with changes in vigilance (Papadelis et al., 2007). However, this feature is not so frequently used as ERP to indicate vigilance.

6. EEG analysis methods

6.1 Signal processing procedure

An EEG-based human-machine-interface, which is also known as brain-computer-interface (BCI), is aimed at translating the brain activities into commands for the computer system. Therefore, after EEG signals recorded, it is of great importance to process these signals into proper commands for controlling the device. As we are interested in estimating the mental states in real time for online support to the system, classic spectral analysis or ERP analysis are normally performed offline on the recorded brain signals of large scale, and are not particularly suitable for this. Whereas, an online analysis is able to provide real-time information on user's mental states and is optimal for this application.

Due to the fact that the online analysis, in contrast to offline analysis, processes data at the same time as recording, it has particularly higher requirements on the analysis. During online analysis, it's vital to take the most relevant data and the processing time of algorithm into account.

This online analysis should enable the translation of EEG data into statements about the mental state of the user. Thus, based on neuroscientific knowledge, specific properties of EEG that correspond to the aspect of cognitive state and a proper algorithmic model should be determined. This model should be capable of not only detecting patterns in the given data, but also in new sets of data from the same type, thus optimizing the prediction of investigated mental state.

Normally, an online analysis consists of preprocessing, feature extraction, and classification (Duda et al., 2000), as illustrated in Figure 6.1. The classification includes classifying on features extracted from the data with given classes, as well as applying the classifier on new feature vectors.

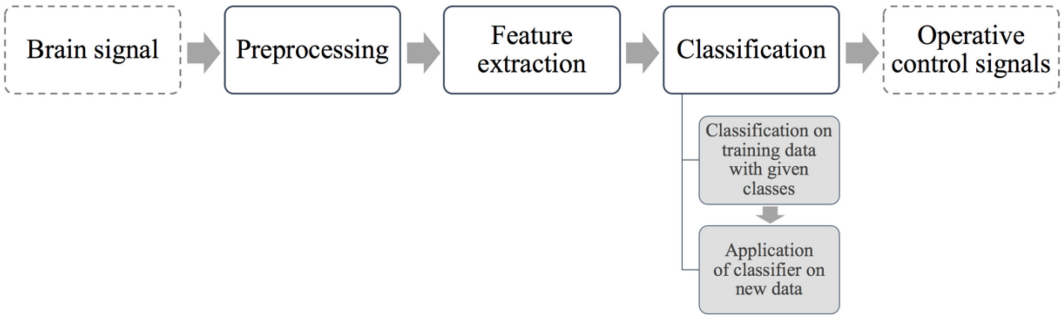


Figure 6.1 Online processing procedures (Duda et al., 2000)

More specifically, in the preprocessing and feature extraction phases, features that are expected to carry information of interest are determined. Then a classification algorithm is fed with feature vectors from the data with labels, i.e. the classes of data should be known. Based on the known classes, the most informative data characteristics are determined and the classification model is trained. Subsequently, the trained model is applied to classify new data in appropriate classes.

The main performance measure in the analysis process is classification accuracy, i.e. the accuracy with which a model, trained on labeled data of an individual, can classify new data with respect to the mental state of the user.

6.2 Preprocessing and feature extraction

6.2.1 Preprocessing

During EEG signal analysis, the first procedure is to reconstruct the data that reflects the most investigated information. This step can be achieved by selecting specific frequency bands or filtering the data along spatial or temporal properties, which can be all categorized as filters.

A filter is a system that processes a discrete signal $\overrightarrow{x_t}$ and converts it into a filtered, modified signal $\overrightarrow{y_t}$. In this context, feature extraction can also be understood as a filter.

$$\overrightarrow{x_t} \rightarrow Filter \rightarrow \overrightarrow{y_t}$$

Spatial filters

Although the EEG allows imaging of the electrical brain activity with a very high temporal resolution, it offers poor spatial differentiation. Due to the nature of the generation of electrical field potentials in the brain and the different electrical properties of the human head, these fields spread differently onto the surface of the skull, leading to spatial distortion or smearing of the EEG signals. In addition, the distance between the electrode positions and the choice of the reference electrode contributes to the blurring of the assignment of special signal sources to the electrode positions (Nunez et al., 1994). Since signals on the electrodes are also contaminated with noise, it is also shown that spatial filters are able to improve the signal-to-noise ratio, since redundant information is removed from the electrode environment.

The proper selection of spatial filters is determined by the location and extent of the control signal and the locations and extents of the various sources of EEG or noise. The most often used spatial filtering methods in EEG data preprocessing are channel selection, common average reference (CAR), and Laplacian filter.

Based on neuroscientific knowledge, cognitive processes are mostly represented in specific EEG channels. Channel selection is a method to reduce the number of channels for use, in order to focus on the most discriminative parts of the EEG data stream and ease the data processing for the classification (Zander, 2012).

EEG data are measured by calculating the electrical potential difference between each recording electrode and a reference electrode. In common average reference, the average value of all electrodes (the common average) is the reference and subtracted from the potential at each electrode (McFarland et al., 1997). This method doesn't need the recording of reference electrode, thus reducing the impact of reference brain activity and problems associated with the actual physical reference (Syam et al., 2017).

If a study focuses on single electrodes that are spatially distant from each other, a Laplacian filter can be used to enhance the signal-to-noise ratio. A Laplacian filter is obtained by subtracting the averaged potential of adjacent electrodes in the 10-20 system from the current electrode potential. It is assumed that noise induced by artifact sources is almost at the same intensity in each electrode, and the intensity of artifact sources is usually higher by some orders of magnitude and distributed by volume conduction. Therefore, a Laplacian filter should enable the elimination of noise while accentuating the localized activity of interest (McFarland et al., 1997).

Temporal filters

Temporal filters transform the relationship of the data within each EEG channel, therefore, they modify the time course of an EEG signal relative to itself (Zander, 2012). The most often used temporal filters are resampling and baseline correction.

Resampling decreases the sampling rate of the recorded signal, by skipping samples equally distributed over time. Resampling can only be applied in offline analysis due to its non-causality. In online studies, the sampling rate should be a fixed value for the whole analysis process.

Baseline correction is most frequently used in ERP analysis, and helps to observe peak amplitudes relative to the basic level of the recorded signal (the baseline). The baseline is the average of samples recorded in the segments, which should only be affected by noise, but not by cognitive processes.

Spectral filters

Spectral filters can also be categorized in temporal filters and are designed for the effects on the spectrum of the signal. Spectral filters are implemented to select the specific EEG sub-frequency bands carrying the neuro signals of interest, since many cognitive states are specifically represented in certain frequency domains. There are several commonly used techniques to define spectral filters, such as fast Fourier transform (FFT), finite impulse response (FIR) filters, and infinite impulse response filters (IIR).

Examples of often used frequency reductions through spectral filters include high-pass filter, low-pass filter, band-pass filter, and Notch filter. High-pass filter preserves all frequencies above the selected threshold, while low-pass filter preserves frequencies below the threshold. Band-pass filter is a combination of the above two filters, and keeps all frequencies between two thresholds. Contrarily, Notch filter retains all frequencies but those between the two selected thresholds.

6.2.2 Feature extraction

Feature extraction is the process to describe the recorded and preprocessed EEG signals using a compact and relevant signal values known as features. These features are represented as a feature vector. Feature extraction is aimed at transforming multichannel EEG data into a significant and reduced dimension feature vector. In machine learning and pattern recognition, analysis with a large number of data requires a large amount of memory and computation power, also it may lead to the overfitting of a classification algorithm to training samples and poor generalization to new samples. Feature extraction is a method of constructing combinations of the variables to get around these problems while still describing the data with sufficient accuracy, facilitating the subsequent pattern recognition and classification of acquired brain patterns to develop operative command signals for the system.

There are a great variety of feature extraction techniques, such as power spectral density (PSD) values (Chiappa & Bengio, 2004; Millan & Mourino, 2003), autoregressive (AR) parameters (Penny et al., 2000; Pfurtscheller et al., 1998), principal component analysis (PCA) (Boye et al., 2008), independent component analysis (ICA), and common spatial patterns (CSP).

Power spectral density (PSD)

Power spectral density is a frequently used feature to identify the correlated neural activity (Unde & Shriram, 2014). It estimates the power distribution of EEG data in a specific frequency range. It is expressed in a Fourier transform of the autocorrelation of recorded EEG signal and reveals the overall frequency content

of particular neural activity. Even using different classifier types, the extracted frequency information through power spectral density can consistently distinguish between different neural activities (Hu et al., 2006).

Autoregressive (AR)

Autoregressive modeling is used to describe certain time-varying processes, and holds the advantage in efficient algorithms for parameter estimation. It is useful for spectral parameter analysis and describing the stochastic behavior of a time series (Schlögl, 2000). In an autoregressive model, the variable of interest is usually predicted by using a linear combination of its previous values, and its coefficients are used as feature vectors.

Principal component analysis (PCA)

Principal component analysis is an orthogonal transformation converting a series of values of possibly correlated variables into a set of data of linearly uncorrelated variables, which are called principal components. PCA sorts the most relevant signal components of the EEG data according to their decreasing variance. This sorting allows the separation of neural responses into different components. The first few most significant components represent most of the signal dynamics, while the rest less significant components are removed. This enhances the signal-noise-ratio by rejecting noise components and reduces the EEG data dimensionality. As a result, PCA can be applied in identifying and rejecting artifact components, and thus reducing the feature vector dimensionality (Boye et al., 2008).

Independent component analysis (ICA)

Independent component analysis is a method for separating a multivariate signal into additive subcomponents, assuming that the subcomponents are independent from each other. Since the different artifacts in EEG data are usually independent of each other, ICA is able to identify the artifacts embedded in obtained EEG data from multiple cognitive activities. It is widely used to remove ocular artifacts in EEG data (Flexer et al., 2005) and attenuate the influence of peripheral muscular activity (Fatourechi et al., 2007).

Common spatial patterns (CSP)

Common spatial pattern and its derivations will be introduced in a more specific way in the following chapter.

6.2.3 Common Spatial Patterns

Common Spatial Patterns (CSP) is a useful tool that can be applied for feature extraction in spectral and temporal domain, i.e. it's a spatio-temporal filter. It is aimed at classifying different conditions that are characterized by the amplitude modulation of brain rhythms (Blankertz et al., 2007), which is captured by spatial filters. This transforms the multichannel EEG data into a sub-domain by maximizing the variations among different classes and minimizing the similarities. The spatially filtered signal should have maximum variance for one class and minimum variance for the other class (Ramoser et al., 2000). This is implemented by maximizing the following function:

$$\mathrm{CSP}(\mathbf{w}) = \frac{\mathbf{w}^{\mathrm{T}}\mathbf{S}_{\mathrm{d}}\mathbf{w}}{\mathbf{w}^{\mathrm{T}}\mathbf{S}_{\mathrm{c}}\mathbf{w}}$$

with

$$\mathbf{S}_{\mathrm{d}} = \mathbf{\Sigma}^{+} - \mathbf{\Sigma}^{-},$$

$$\mathbf{S}_{\mathrm{c}} = \mathbf{\Sigma}^{+} + \mathbf{\Sigma}^{-}.$$

Here $\mathbf{\Sigma}^{+}$ and $\mathbf{\Sigma}^{-}$ represents estimates of covariance matrices of the bandpass filtered EEG signals of the two conditions, and $\mathbf{w}$ represents a spatial filter. The maximum of function CSP($\mathbf{w}$) provides the spatially filtered EEG signals with maximum band power difference among distinct classes (Blankertz et al., 2007). Figure 6.2 presents a typical distribution of CSP patterns in the motor imagery study, where patterns illustrate how the presumed sources project to the scalp.

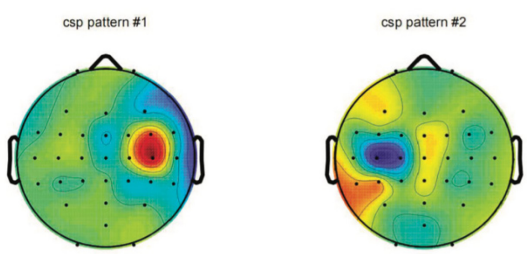


Figure 6.2 A typical pair of CSP patterns in the motor imagery study. The pattern on the left hand side shows weights that are optimized for imagined left hand movements. The other pattern is optimized for right hand movements. Red indicates positive weights, while negative weights are in blue. (Zander, 2012)

The CSP holds a number of advantages. It is computational efficient, easy to implement, possessed of high classification performances, and can be applied in versatile contexts. However, the performance of CSP greatly depends on its operational frequency band. To solve the problem of manually selecting the operational frequency band, there exist a bunch of derivations of CSP, including Common Spatio-Spectral Pattern (CSSP), Common Sparse Spectral Pattern (CSSSP), Filter Bank Common Spatial Pattern (FBCSP) and so on.

The Common Spatio-Spectral Pattern (CSSP) optimizes a simple filter that employs a one time- delayed sample with the CSP algorithm (Lemm et al., 2005). The Common Sparse Spectral Spatial Pattern (CSSSP) improves the CSSP algorithm by simultaneous optimizing an arbitrary FIR filter within the CSP algorithm (Dornhege et al., 2006).

The Filter Bank Common Spatial Pattern (FBCSP) consists of four stages, including frequency filtering, spatial filtering, feature selection and classification (see Figure 6.3). In the first stage, EEG data are bandpass filtered into multiple frequency bands. In the second stage, CSP features are extracted from each of the frequency bands. In the third stage, a feature selection algorithm is applied to automatically select discriminant pairs of frequency bands and corresponding features. In the fourth stage, a classification algorithm is used to classify the CSP features (Ang et al., 2008).

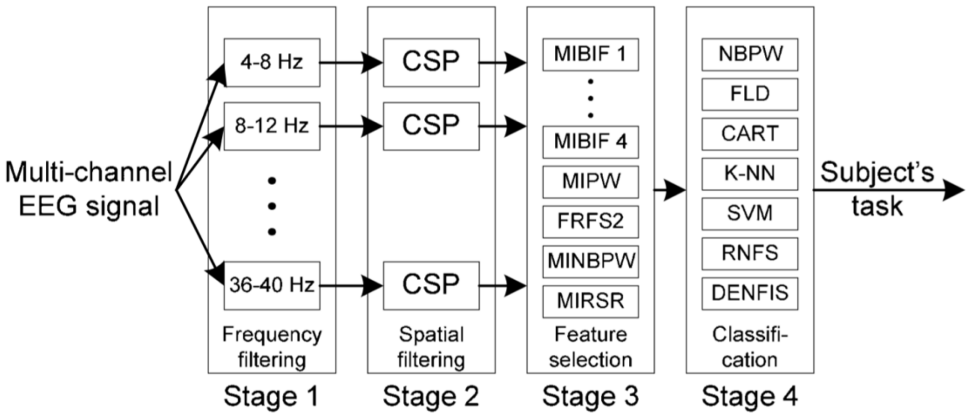


Figure 6.3 Architecture of the Filter Bank Common Spatial Pattern (FBCSP) approach (Ang et al., 2008)

In comparison to other derivate algorithms, the FBCSP added the employment of a feature selection algorithm to automatically select discriminating CSP features. Hence, FBCSP is more efficient since it deploys only those effective spatial filters whose pairs of CSP features are selected, thus significantly reducing the computational complexity. For this reason, the FBCSP was applied in our studies.

6.3 Classification

6.3.1 Different classifiers

After the feature extraction step, classification follows up and is used to translate the extracted feature sets into operative commands and assign set of extracted features with appropriate class labels. Classifiers are aimed at separating the feature projections from one another. The processing time of the classification system is greatly influenced by the size of extracted features and the complexity of classifiers. Generally, classification algorithms can be categorized into linear and nonlinear classifiers.

Linear classifiers

Linear classifiers are in principle an establishment of a linear relationship/function between input and output variables of a classification system, in order to categorize distinct classes of brain signals. Examples of linear classifiers are Linear Discriminant Analysis (LDA) and Support Vector Machines (SVMs).

Fisher developed the LDA technique in 1936, which follows the principle of linear separability of two class problems. It establishes a linear mathematical function known as hyperplane to separate the distinct brain signals from the extracted features (Bansal & Mahajan, 2019). LDA has been widely used in many EEG applications. In our studies, LDA was implemented as the classifier, and a more detailed introduction to LDA will be given in the following part (Chapter 6.3.2).

Support Vector Machines are implemented to estimate one or more decision boundaries so as to maximize the margin between nearest training support vectors and the decision boundaries (Xiang et al., 2007). Support Vector Machines do not estimate the full covariance matrices for both classes, and they are thus more robust for a limited number of training data. It's also possible to apply an SVM in multiclass nonlinear classification, in combination with the kernel trick to raise the classification accuracy.

Nonlinear classifiers

Nonlinear classifiers are usually used when any algorithmic solution among input and output variables are not suitable for the classification (Bansal & Mahajan, 2019). The frequently used algorithms include artificial neural networks (ANNs), k-nearest neighbors, and SVMs (can also be linear).

The ANNs are able to learn directly from training data and then classify the input data accordingly. Usually, an ANN consists of an input layer, possibly one or several hidden layers and an output layer, each with one or more neurons. Each neuron's input is based on the output of the previous layer's neurons. In order to

minimize the difference between target and actual output, ANNs develop training algorithms to adjust weights on input and hidden layer neurons. Among all ANNs, the multilayer perceptron (MLP) is the most widely used ANN in classifying multiclass neural activities (Lotte et al., 2007).

Based on the principle that different class EEG features normally constitute separate clusters during feature mapping in feature space, k-nearest neighbors is implemented to assign to an unseen point the most dominant class within the training set. Moreover, very close neighbors are supposed to belong to the same class. The input feature vector can be efficiently classified according to its distance to the neighbor in a multiclass environment (Kayikcioglu and Aydemir, 2010).

There are many other nonlinear classifiers, such as Bayes quadratic, Hidden Markov model, Mahalanobis distance. However, these classifiers are not so widespread as the above-mentioned techniques.

6.3.2 Linear Discriminant Analysis

For the selection of classification technique, a series of factors should be taken into account.

Müller et al. (2003) discussed on classification approaches in the area of brain-computer interfaces and stated so: "…overall it was agreed that simplicity is generally best and therefore, the use of linear methods is recommended wherever possible." (Müller et al., 2003).

More complex classification methods, such as the non-linear Support Vector Machines (SVM) appear to be unsuitable for classifying patterns in EEG data, since the necessary recalculations of the classifier would be too time-consuming.

Moreover, nonlinear classifiers are sensitive to overtraining and overfitting. The study of Boostani and Moradi (2004) proves that LDA outperforms nonlinear classifiers, because nonlinear classifiers boost mislabels, which are very likely to occur in such noisy and non-stationary data as EEG. In contrast, Linear Discriminant Analysis (LDA) is optimal in the sense that it minimizes the risk of misclassification for new data from the same distributions (Duda et al., 2001). Simpler approaches are significantly more robust against interferences in the EEG signals.

LDA is easy to implement and computationally cheap. Besides, it shows impressive results that are comparable with more complex classification methods (Blankertz et al., 2011). Overall, with its simple, robust, non-overfitting and computational requirements, LDA outperforms other classifiers and is most suitable for the application in this dissertation.

The core of the linear discriminant analysis is a discriminant function. This discriminant function **y** can be described as a linear function of the vector **x**:

$$\mathbf{y} = \mathbf{w}^{\mathrm{T}}\mathbf{x} + \mathbf{w}_0,$$

where **x** represents input EEG feature vector, **w** is the weight vector, and $\mathbf{w}_0$ is the threshold to define decision boundaries. The weight vector **w** is defined as

$$\mathbf{w} = \hat{\boldsymbol{\Sigma}}_c^{-1}(\hat{\boldsymbol{\mu}}_2 - \hat{\boldsymbol{\mu}}_1)$$

where $\hat{\mu}_i$ is the estimated mean of the class i and $\hat{\Sigma}_c = 1/2(\hat{\Sigma}_1 + \hat{\Sigma}_2)$ is the estimated common covariance matrix.

An assignment to two different classes is made by evaluating this function on the elements $\vec{x}$ of the feature space, and the distinction between two classes depends on the sign of this function $y(\vec{x})$:

$$\mathbf{w}^{\mathrm{T}}\mathbf{x} + \mathbf{w}_0 \geq 0 \rightarrow \text{class 1}$$

$$\mathbf{w}^{\mathrm{T}}\mathbf{x} + \mathbf{w}_0 < 0 \rightarrow \text{class 2.}$$

This makes it clear that the differentiation of different features in a multi-dimensional feature space can be described by solving a one-dimensional problem. For this purpose, the discriminant function projects all elements of the feature space into the one-dimensional space **y**. For a two-dimensional feature space, the separating decision level is simply a straight line.

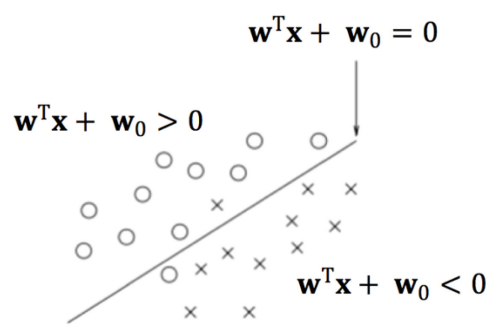


Figure 6.4 A hyperplane separating two classes. The crosses are one class, and the circles are the other class.

In the course of this work the method described by Blankertz et al. (2011) used the form of LDA according to Friedman (1989). This approach was implemented in MATLAB through the work of Kothe (2009).

6.4 Method applied in this work

Overall, in the studies of this work, filter bank common spatial patterns (FBCSP) was applied to extract features relating to power in theta (4-7 Hz) and alpha (8-13 Hz) bands, which are highly related to mental workload. Linear discriminant analysis (LDA) was used to separate the classes. In the testing session, classifier outputs were calculated by applying the calibrated classifiers to testing data, and difference of classifier outputs of two classes were then compared by means of permutation tests.

The more detailed method and comparison with some other methods will be explained in Chapter 7.6.

7. Experiment 1 – Validation of task-independent workload classifier with EEG[9]

7.1 Introduction and hypothesis

As discussed in Chapter 3.2, workload is a crucial factor influencing driver's performance in the take-over situations, and EEG has been proven as an effective tool for workload detection. Current approaches to workload detection with EEG are typically limited to the investigation of a specific task. That means, for each task it is necessary to calibrate a single classifier which could provide highly accurate classifications of cognitive state. However, it is impractical in operational environments since it is a cumbersome and time-consuming procedure. Therefore, we are motivated to investigate whether we could calibrate a classifier that is transferable between different tasks. In the present study, the goal was to develop a method for quantifying mental workload that generalized across tasks.

This study was aimed at investigating the task-independency of a classifier to indicate actual workload level using EEG, which is based on the preliminary study by Krol and his colleagues (2017). We extended the application of this task-independent classifier not only to tasks that are commonly used in workload research but also to application-oriented tasks.

Therefore, we hypothesize that,

- it is possible to calibrate a task-independent classifier to identify whether a user is under heavy/low workload across different activities.

7.2 Introduction of the tasks

In this study, all participants need to perform a calibration task, and five testing tasks of two different workload levels each. The five testing tasks were: n-back task (n-back), backward digit span task (span), addition task (add), scrambled word recovery task (word) and mental rotation task (rotation), each consisting of two difficulty levels to induce different workload. N-back task (Kirchner, 1958) and digit span task

[9] This chapter is based on the following conference paper: Zhang, X., Krol, L. R., & Zander, T. O. (2018, October). Towards Task-Independent Workload Classification: Shifting from Binary to Continuous Classification. In *2018 IEEE International Conference on Systems, Man, and Cybernetics (SMC)* (pp. 556-561). IEEE.

(Turner & Engle, 1989) are commonly used for inducing workload and base on two different theories of mental workload (Redick & Lindsey, 2013). The addition, word recovery, and mental rotation tasks are more practical tasks representing different aspects of cognitive abilities, i.e. mathematical, linguistic and spatial ability, and could be used to investigate the applicability of workload detection in more real-life tasks (e.g. Walter et al., 2017; Koshino et al., 2005; So et al., 2017). In the following, an introduction to the calibration task and the five testing tasks, as well as their settings of two different task difficulty levels will be given.

Calibration task: This task consisted of 40 trials in random order (20 with high, 20 with no workload). During high workload conditions, participants were presented with an equation of the form a – b, instructing them to count backwards from a in steps of b (i.e. a modified Brown-Peterson distraction technique). a was any integer between 200 and 1200; b ranged from 6 to 19, excluding 10 and 15. During no-workload phases, a crosshair was shown in the middle of screen and participants were instructed to relax, with eyes open. In 50% random trials of both high and no workload conditions, there were 10 small visual lightly-colored "sparkles" wandering smoothly over the screen in random walks governed by perlin noise, providing visual distraction. This 50% chance of sparkles was presented so as to evenly balance eye movements between classes. A self-paced break was implemented after every 10 trials. Each trial lasted 10 seconds, for a total of 200 seconds of EEG data per workload class.

N-back: Participants were given a sequence of numbers and had to indicate when the currently displayed number was equal to the number displayed n elements previously. Participants performed 24 trials of 25 seconds each, among them 12 trials with high difficulty, 12 with low difficulty. We used 1-back (low difficulty) and 3-back (high difficulty) varieties of the n-back task to manipulate the task difficulty and the task load. The sequence of high/low workload trials is randomized. The same applies to the following tasks.

Span: Participants were given a sequence of numbers which they then had to repeat in reverse order. This task consisted of 2-digit span task (low difficulty) and 6-digit span task (high difficulty) conditions, and had 20 blocks each lasting 30s: randomly 10 with high difficulty and 10 with low difficulty (the same for the following three tasks).

Add: Participants were given two numbers to add. The workload level was varied according to the Q-value of the equation (Walter et al., 2017; Thomas, 1963): for low difficulty 2-digit additions with Q-value from 2 to 2.5, for high difficulty 4-digit additions with Q-value from 4 to 5.

Word: Participants were presented scrambled German words and they had to recover the original words. The workload level was differentiated by the use frequency and syllable number of the words (Brysbaert et

al., 2011). The scrambled words in the low-difficulty condition are frequently-used 2-syllabled words, and those in high-difficulty condition are infrequent 3-syllabled words.

Rotation: Participants were presented either two figures of 2D objects containing 6 squares (low difficulty) or two figures of 3D objects containing 9 cubes (high difficulty) and were instructed to determine whether they were the same object (dataset from So et al., 2017).

7.3 Pre-study: Manipulating workload of different tasks

In order to validate the difference between two difficulty levels in each of these five testing tasks, a pretest with 9 participants reporting mental workload levels was conducted. It is assumed that the subjective ratings of mental workload should be significantly different between the two difficulty setting levels of each task.

9 participants whose age ranged from 21 to 31 (M=25.67, S=3.20) participated in this pre-study. The experiment took place in a well-controlled laboratory, with a 25-inch screen standing on a table in front of the participants. Subjective ratings on perceived mental workload of all tasks were recorded by means of the NASA-TLX questionnaire (German version from Unema et al., 1988).

During the experiment, the participants were firstly asked to read the instruction of this experiment, illustrating all different types of task. Before performing each task, they had a few minutes to practice and get familiar with all tasks. Following that, participants were instructed to perform successively all the five tasks consisting of two different difficulty levels. In each task, there were several trials of the two difficulty levels each. The sequence of five tasks was randomized, so was the sequence of trials of the two difficulty levels. After each task session, participants reported their subjective workload on both difficulty levels, by answering the NASA-TLX questionnaire.

Mean values and standard deviations of NASA-TLX ratings across all participants in each task in both difficulty conditions were calculated. T-tests were computed for each difficulty pairs of five tasks and their statistical significance were also examined, by means of SPSS Statistics (version 26.0.0.0).

Table 7.1 lists the mean scores and standard deviations of NASA-TLX ratings across all participants in each task for two difficulty levels, as well as the p-values of t-test between two difficulty levels in each task. Figure 7.1 shows the NASA-TLX ratings under all conditions in a more direct way.

Table 7.1 Average ratings of NASA-TLX across all participants and comparison results between two difficulty levels in each task

	N-back		**Span**		**Add**		**Word**		**Rotation**	
	Low	**High**	**Low**	**High**	**Low**	**High**	**Low**	**High**	**Low**	**High**
M	14.23	42.56	13.36	44.01	22.30	47.31	20.37	45.26	20.28	37.93
SD	6.85	17.43	6.97	10.26	12.36	13.02	11.69	11.43	10.33	9.83
p-value	<0.001***		<0.001***		<0.001***		<0.001***		<0.001***	

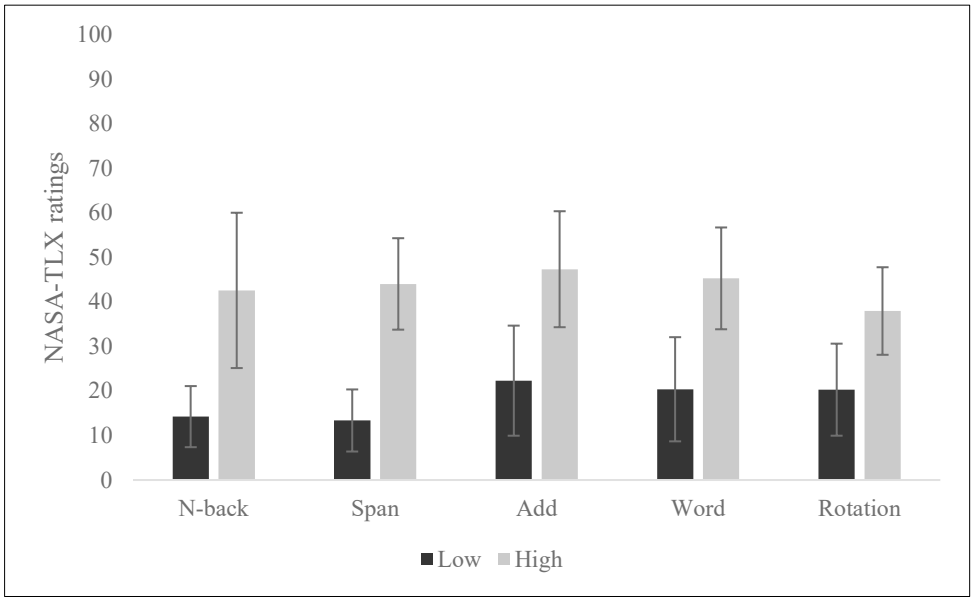


Figure 7.1 Average NASA-TLX ratings across all participants in each task

From the results above, we could see significant differences between two difficulty levels in all five tasks. Thus, it can be concluded that the two difficulty levels of each task are differentiable, and they could further serve as the "ground truth" for the classification later when classifying EEG signals between different mental workload levels.

7.4 Methods

Based on the results from pre-study, this experiment was conducted to investigate a task-independent classifier in mental workload classification.

7.4.1 Participants

This experiment was conducted in a well-controlled laboratory and 15 participants (8 females and 7 males) between the ages of 24-35 (M=28.73, SD=4.17) participated in this experiment. All participants reported to be free of illness and medication. Caffeine, tobacco, and alcohol were prohibited prior to participating in the experiments. Participants were rewarded either with a cash payment (10 Euros per hour) or with a certification of student experimental hours for their participation.

7.4.2 Experimental design

This study featured a one-factorial within-subject design. The independent variable here is the workload level of different tasks, i.e. the different task difficulties of each task. Each task consists of two levels (low and high) and their differences were examined in the pre-study (see Chapter 7.3) by measuring their subjective workload levels with NASA-TLX. The dependent variables here are performance of classifier, i.e. the classification accuracies, as well as subjective ratings of task workload based on NASA-TLX.

7.4.3 Procedure

Firstly, participants were provided with a general introduction to the experimental procedure and an informed consent form, and if no disagreement, they signed this form. Following this, participants were asked to fill out a demographic questionnaire. After that, the EEG electrode cap was placed on the participant and conductive gel was applied so that impedance of electrodes was adjusted for proper use. After all these preparation work, participants were seated in front of a computer screen and began the real experiment phase.

This experiment consists of a training session and a testing session. In the training session, a calibration task was performed, which has been described concretely in Chapter 7.2. In the testing phase, there were five tasks, each with two different task difficulty levels. The sequence of these five task sessions was randomized, so was the sequence of trials of both task difficulty levels. After the first and last trials of each difficulty level, participants reported their subjective mental workload ratings with NASA-TLX (German version). After finishing all the tasks, participants could remove the electrode cap and wash their hairs. The electrode cap was also cleaned up by the conductor.

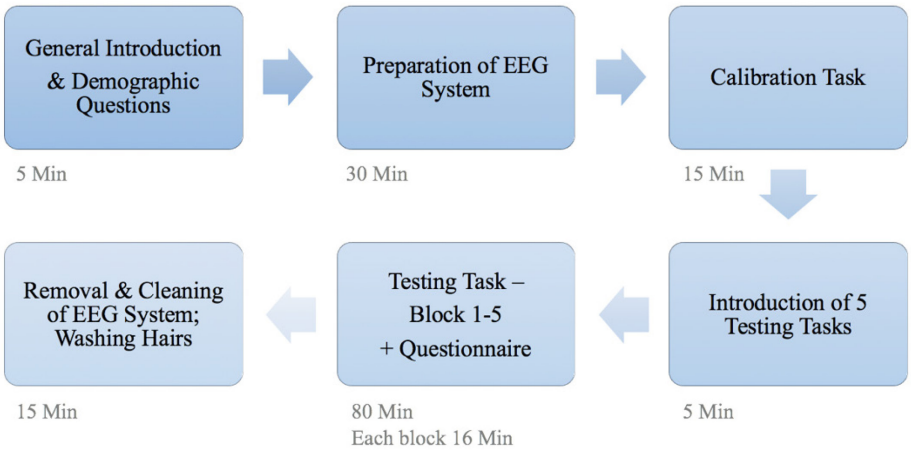


Figure 7.2 General procedure of Experiment 1

7.4.4 Experiment apparatus

The participants were instructed to carry out the calibration task and testing tasks on a 25-inch screen standing on a table in front of them. EEG data were recorded continuously using a 64 active Ag/AgCl electrode mounted according to the extended 10–20 system on an elastic cap (ActiCap by Brain Products GmbH, Gilching, Germany). The signal was sampled at 500 Hz and amplified using BrainAmp DC amplifiers (Brain Products GmbH, Gilching, Germany). All electrodes were referenced to FCz and the ground electrode was placed at position AFz.

7.4.5 Data analysis

Subjective ratings based on NASA-TLX were firstly calculated and their differences between two workload conditions of each task were then analyzed with t-test, in order to verify the differentiability of these two conditions for each task.

On the basis of subjective data analysis, further exploration on classification accuracy was executed. To investigate the task-independence of the workload classifier, individual classifiers were first calibrated based on data recorded in the calibration task. The data were divided into consecutive 1-second epochs of high versus low condition data. Filter bank common spatial patterns (FBCSP) was applied to extract features relating to power in theta (4-7 Hz) and alpha (8-13 Hz) bands with three patterns per band. Linear discriminant analysis (LDA) was used to separate the classes with a 5-fold nested cross-validation with margins of 5. The calibrated classifiers were then applied to 1-second epochs taken from the five testing tasks. Classification was computed between two different workload levels (high vs. low workload).

Furthermore, the classifier outputs of two workload conditions in all five tasks were calculated. Differences were also compared by means of permutation tests.

Calibration and classification was done using the open-source MATLAB-based toolbox BCILAB (version 1.2) (Kothe & Makeig, 2013).

7.5 Results

7.5.1 Subjective measures

Table 7.2 lists the means and standard deviations of workload ratings and p-values from the t-test between two conditions in each task. Figure 7.3 shows the mean subjective ratings of both conditions of all five testing tasks. From these we could observe that, the overall workload ratings increased in line with task difficulty, showing higher ratings for the more difficult conditions. The results of t-test also demonstrated significant differences in subjectively perceived mental workload between two conditions in each task.

Table 7.2 Average ratings of NASA-TLX across all participants and comparison results between two difficulty levels in each task

	N-back		**Span**		**Add**		**Word**		**Rotation**	
	Low	**High**	**Low**	**High**	**Low**	**High**	**Low**	**High**	**Low**	**High**
M	16.00	50.50	12.33	51.00	15.00	56.50	22.50	59.33	22.17	42.67
SD	14.72	20.18	11.93	25.09	9.64	24.12	12.61	18.41	9.95	18.81
p-value	<0.001***		<0.001***		<0.001***		<0.001***		<0.001***	

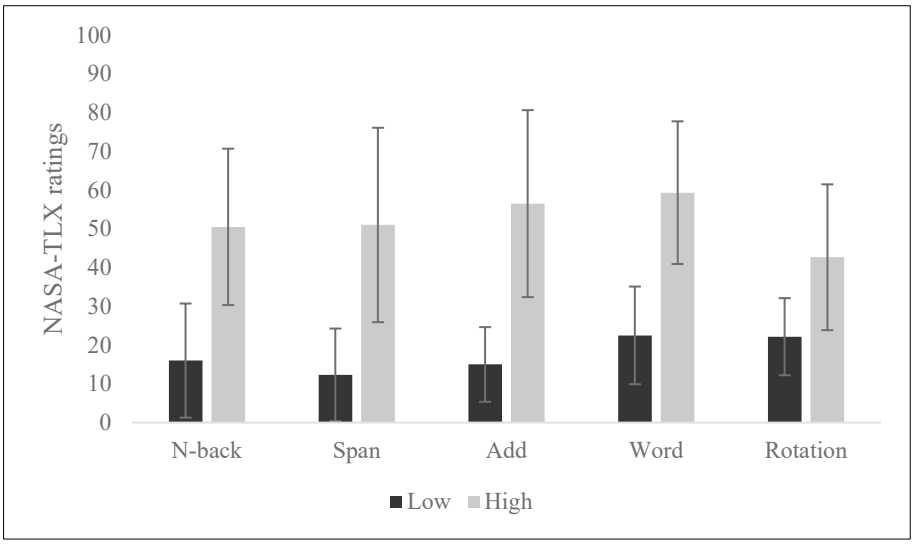


Figure 7.3 Average NASA-TLX ratings across all participants in each task

7.5.2 Performance of task-independent workload classifier

Table 7.3 lists the classification accuracies of a classifier trained on data from the calibration task, and applied to the data of the five testing tasks. For each testing task, the accuracy is calculated from classification between the high and low workload conditions of this task. The mean estimated classification accuracy over participants in the calibration task reaches 76%±11 (significance is reached at 55% or above). Applied to the five testing tasks, however, no meaningful performance can be reported.

Table 7.3 Classification accuracies of High vs. Low workload conditions

Participant	Calibration	Classification Accuracy for Each Testing Task between High and Low Levels (%)				
		N-back	Span	Add	Word	Rotat.
1	64.25	50.42	51.09	56.22	56.06	54.33
2	83.50	64.79	60.47	61.84	59.87	53.48
3	67.00	60.42	60.31	62.30	60.66	53.91
4	60.00	53.99	64.04	55.20	50.79	54.01
5	78.25	53.54	48.75	55.02	59.12	52.09
6	71.25	55.21	56.27	54.47	57.94	56.24
7	92.00	49.79	61.52	64.33	64.50	49.17
8	78.00	57.50	65.21	61.26	54.56	49.42
9	91.50	59.38	66.83	51.13	53.05	57.49
10	82.25	53.54	60.34	57.28	49.67	54.18
11	88.25	60.21	67.84	63.96	50.78	49.66
12	81.75	59.17	50.84	73.72	61.87	58.16
13	79.50	53.33	51.24	60.40	55.01	53.48
14	62.25	49.17	58.01	53.77	46.65	49.03
15	60.50	57.29	54.82	56.12	55.82	56.54
M	76.02	55.85	58.51	59.13	55.76	53.41
SD	11.09	4.47	6.16	5.70	5.00	3.02

Table 7.4 lists the mean classifier output values produced when applying the calibrated classifier to the data of the five testing tasks as well as the no-workload baseline recording. Values could vary between 0 and 1, and represent the classifier's predictions, with 0.5 representing complete uncertainty between the classes, 0 representing certainty for class I – low workload (i.e. the classifier predicts a 100% chance of the data belonging to class I, low workload) and 1 certainty for class II (high workload).

Table 7.4 Mean classifier outputs of each tasks in High and Low workload conditions

Participant	Mean Continuous Classifier Output									
	N-back		Span		Add		Word		Mental Rotation	
	High	Low	High	Low	High	Low	High	Low	High	Low
1	0.31	0.32	0.54	0.52	0.82	0.69	0.76	0.65	0.80	0.69
2	0.71	0.50	0.72	0.56	0.87	0.68	0.77	0.67	0.73	0.67
3	0.49	0.46	0.62	0.41	0.57	0.49	0.64	0.63	0.45	0.38
4	0.56	0.41	0.52	0.39	0.95	0.67	0.89	0.69	0.95	0.86
5	0.41	0.39	0.48	0.52	0.93	0.79	0.75	0.62	0.92	0.92
6	0.47	0.40	0.49	0.36	0.54	0.48	0.60	0.49	0.63	0.54
7	0.66	0.69	0.71	0.57	0.77	0.58	0.60	0.46	0.75	0.79
8	0.44	0.32	0.64	0.43	0.83	0.63	0.83	0.76	0.84	0.84
9	0.54	0.42	0.71	0.53	0.85	0.79	0.77	0.71	0.84	0.72
10	0.28	0.22	0.49	0.30	0.64	0.54	0.73	0.76	0.65	0.61
11	0.35	0.19	0.51	0.23	0.81	0.58	0.87	0.88	0.80	0.79
12	0.44	0.25	0.43	0.41	0.86	0.48	0.78	0.64	0.79	0.65
13	0.37	0.29	0.51	0.50	0.86	0.73	0.64	0.57	0.76	0.70
14	0.37	0.38	0.67	0.47	0.56	0.47	0.49	0.55	0.62	0.64
15	0.53	0.40	0.49	0.43	0.63	0.54	0.48	0.36	0.56	0.49
M	0.46	0.38	0.57	0.44	0.77	0.61	0.71	0.63	0.74	0.68
SD	0.12	0.12	0.10	0.10	0.14	0.11	0.13	0.13	0.14	0.15

Figure 7.4 presents the mean classifier output across all participants. We see that the mean output for the n-back task in both high and low workload conditions are below the middle line (0.5), and those of the addition task, word recovery task, and mental rotation task in are similarly all above 0.5. Table 7.5 lists the results of permutation tests comparing the output within each participant between high and low workload conditions. For each task, over two-thirds of the participants had significantly different classifier output between conditions. Over all participants, in 76% of all tasks the differences between high and low conditions were significant.

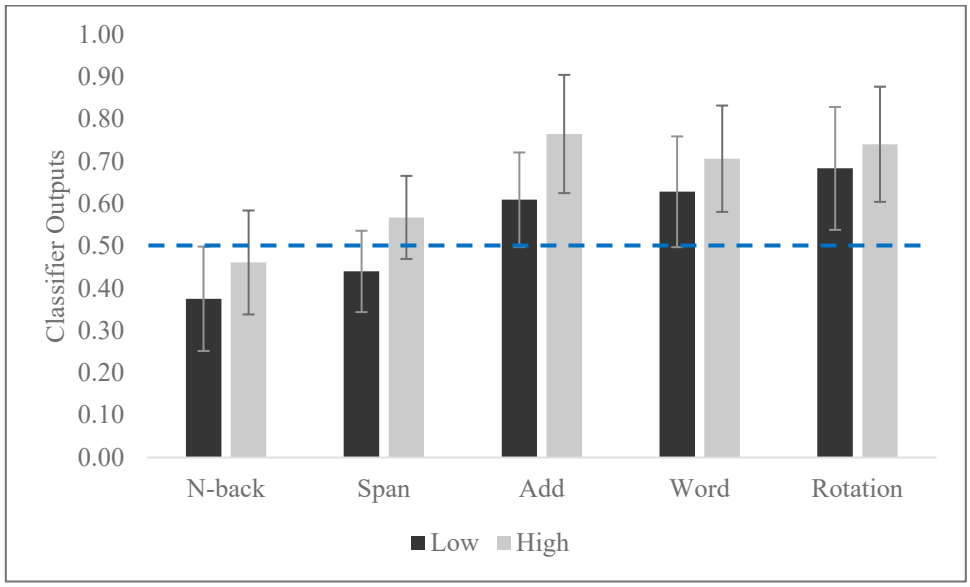


Figure 7.4 Averaged classifier outputs across participants

Table 7.5 Results of permutation tests on outputs of High vs. Low workload conditions of each participant

Participants	P-value of Permutation Test					Percentage of Significant Tasks (%)
	N-back	Span	Addition	Word	Rotation	
1	p = 0.801	p = 0.491	p < 0.001***[a]	p < 0.001***	p < 0.001***	60
2	p < 0.001***	p < 0.001***	p < 0.001***	p < 0.001***	p = 0.001***	100
3	p = 0.002**	p < 0.001***	p < 0.001***	p < 0.001***	p < 0.001***	100
4	p = 0.214	p < 0.001***	p = 0.002**	p = 0.649	p = 0.004**	60
5	p = 0.389	p = 0.191	p < 0.001***	p < 0.001***	p = 0.846	40
6	p = 0.009**	p < 0.001***	p = 0.029*	p < 0.001***	p < 0.001***	100
7	p = 0.239	p < 0.001***	p < 0.001***	p < 0.001***	p = 0.014*	80
8	p < 0.001***	p < 0.001***	p < 0.001***	p = 0.002**	p = 0.762	80
9	p < 0.001***	p < 0.001***	p = 0.002**	p = 0.016*	p < 0.001***	100
10	p = 0.012*	p < 0.001***	p < 0.001***	p = 0.313	p = 0.102	60
11	p < 0.001***	p < 0.001***	p < 0.001***	p = 0.433	p = 0.210	60
12	p < 0.001***	p = 0.347	p < 0.001***	p < 0.001***	p < 0.001***	80
13	p = 0.002**	p = 0.726	p < 0.001***	p < 0.001***	p = 0.004**	80
14	p = 0.634	p < 0.001***	p = 0.015*	p = 0.075	p = 0.604	40
15	p < 0.001***	p = 0.023*	p = 0.001**	p < 0.001***	p = 0.004**	100
Percent. of particip. w. significance (%)	66.7	73.3	100	73.3	66.7	Average: 76

[a] * p < 0.05, ** p < 0.01, *** p < 0.

(p-value highlighted with grey means insignificant)

7.5.3 CSP pattern analysis

Figure 7.5 depicts the first spatial pattern pairs of both theta and alpha bands, and reflects an increase in the theta power at frontal electrodes and a decrease in alpha power at parietal electrodes, under a higher mental workload condition. For the theta band, we could observe a frontal area in pattern 1, which is more yellow compared to pattern 2, under the high workload condition, and this indicates an increase in the theta power in frontal area. Similarly, for the alpha band, higher workload reflects a more blue area in the parietal area in pattern 1 compared to pattern 2, which means there is a decrease in alpha power in parietal area.

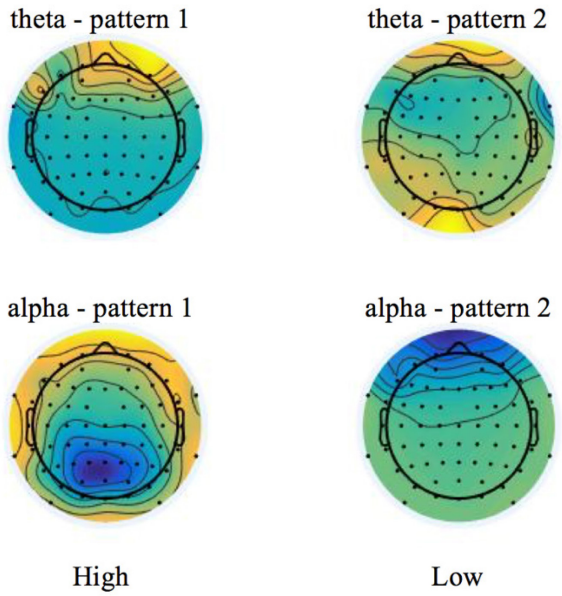


Figure 7.5 First CSP pattern pairs of two frequency bands: theta band (4-7 Hz) in the first row and alpha band (8-13 Hz) in the second row (blue means negative, yellow means positive)

7.6 Discussion

The main aim of this study was to investigate the feasibility of a task-independent classifier for mental workload detection. Such a classifier could potentially greatly simplify the calibration session in more practical environments. To manipulate workload levels, we exposed participants to one no-workload session and five different tasks inducing mental workload, each containing two different workload levels.

From Table 7.3 it can be concluded that a direct transfer between tasks is not sufficient to separate high and low workload levels within new tasks with acceptable accuracy. This, however, was to be expected: different tasks are subjectively more or less demanding, and a direct comparison between “high” and “low” load levels across tasks is thus unrealistic. Instead, we turned to classifier output on a continuous scale.

We can see from Figure 7.4 that the mean classifier output differs for each of the five testing tasks. The mean classifier outputs of both conditions in the n-back task are below 0.5, whereas those in addition, word recovery, and mental rotation tasks are all above 0.5. As the default separating threshold of high versus low workload conditions in our calibrated classifier is 0.5, the unreliable classification results in these tasks can be explained to some extent. Furthermore, the permutation tests on outputs of each subject reveal that for most subjects in most tasks there are significant differences between conditions. This implies that the calibrated classifier did in fact produce reliable differences between the workload levels, but that these differences were simply not centered on the original threshold.

We conclude that, due to inherent differences between different tasks with respect to the induced workload, it is not advisable to directly transfer a binary classifier, calibrated on one task, to another task. Instead, a more reliable measure may be to use the raw classifier output mapped onto a linear scale.

Alternatively, it is possible to adjust the decision threshold for each task individually. However, this would essentially again require a calibration phase for each task specifically, no longer making the procedure task-independent.

We calibrated a classifier on a continuous subtraction task, and applied it to five other tasks that induced two different levels of workload, as well as to a no-workload condition. We found significant differences in the classifier output between conditions. The tasks we used represent a variety of different types of workload. These initial results imply that it is possible to calibrate a single classifier on a standardized workload-inducing task, and use this same classifier to obtain meaningful information concerning workload levels induced by other tasks. This could greatly simplify the implementation of real-time, continuous workload detection into real-world applications.

In comparison to classic mental workload classification methods or other methods investigated in previous studies (e.g. Lei, 2011), the method applied in this work doesn’t limit calibration task to the same one as in testing session. Whereas, we only need to calibrate on a simple, standard workload task and test on other tasks. This method is universal to apply in different tasks and different scenarios, which is much more convenient than other methods.

In future work, we will analyze how the classifier output correlates to the subjective ratings that participants gave after each block. Different participants could have different abilities and different perceptions of tasks, and this may help explain the variance still present in the results. We will also investigate the generalizability of this classifier between subjects, i.e. we will take further steps towards a *universal* workload classifier—one that is both task- and subject-independent.

7.7 Summary

This work was aimed at investigating a task-independent workload classifier, which can be applied to classify mental workload induced by various tasks (including n-back, backward span, addition, word recovery and mental rotation).

Results showed that this task-independent workload classifier, which was calibrated on a standardized workload-inducing task, was able to apply to other tasks and obtain meaningful information concerning workload levels induced by other tasks. Furthermore, this approach is not limited to binary classification of workload but can discriminate it on a continuous metric.

This task-independent workload classifier could greatly simplify the calibration process of different tasks and the implementation of real-time, continuous workload detection into real-world applications.

8. Experiment 2 – Detection of driver's mental workload and vigilance with EEG in highly automated driving

8.1 Introduction and hypothesis

In highly automated driving, driver can safely turn their attention away from the driving tasks, such as texting or watching a movie, but must still be prepared to intervene within some limited time, when called upon by the vehicle to do so. For instance, a car is driving autonomously on a highway and the driver is undertaking some driving-unrelated activity. At some certain time, the car will drive off the highway and the driver needs to take over the control of the car (assuming this L3 autonomous car is only allowed to use its autonomous driving function on highways). Monitoring driver's mental states while the car drives autonomously and before the take-over scenario could help ensure a safer transition from automation to manual driving.

Hence, in this study we investigated the detection of mental workload and vigilance with EEG in the highly automated driving situation. As introduced in Chapter 5, EEG components could reflect variations in mental workload and vigilance level. With the aid of EEG data analysis, it is possible to monitor mental workload and vigilance level non-intrusively and in real time. In this study, we applied a simplified workload classifier based on EEG, whose feasibility has been verified in Experiment 1 (see Chapter 7), to detect mental workload levels of different non-driving tasks. Meanwhile, under different workload conditions, we also utilized an auditory oddball task to indicate driver's vigilance performance while undertaking non-driving activities. Under this circumstance, as higher workload may have negative effect on vigilance and lead to inattention and deteriorated performance, we also investigated workload's influence on vigilance level, i.e. performance of the auditory oddball task.

In brief, in this study we aim at

- investigating whether it is possible to classify between different mental workload levels with EEG and apply the task-independent classifier (which has been proven in Experiment 1) in identifying whether the driver is under heavy or low workload across different non-driving activities (this part will be verified in details in Experiment 2a) in Chapter 9);
- detecting whether and to what extent the target auditory stimuli are perceived, among other auditory distractors, in order to identify the vigilance level under different workload conditions (this part will be thoroughly investigated in Experiment 2b) in Chapter 10);

- verifying that higher workload induces a decrease in vigilance and thus deteriorated performance in detecting targets in the auditory oddball task (this part will also be examined in Experiment 2b) in Chapter 10).

8.2 Introduction of the tasks

This study took place in a simulated autonomous driving scenario (Level 3 automation, i.e. drivers were allowed to have hands and eyes off), simulating the situation before the take-over control. All participants of this study performed simultaneously two tasks: a non-driving task inducing mental workload (primary task), and the auditory oddball task (secondary task). As background, all participants were seated in a driving simulator in the autonomous driving mode, which means, during the whole process of this study, the vehicle is driving autonomously and participants don't need to take any action to drive.

8.2.1 Primary task

The primary task here is to induce different mental workload levels. There are 7 different kinds of non-driving tasks: listening to quiet music (music), listening to news (news), listening to two simultaneously presented voice recordings (talk), watching movies (movie), reading English articles and counting the occurrence of a certain letter (letter), reading English articles and counting the number of adjectives (adjective), and mental calculation (calculation) (the words in the brackets represent these 7 tasks as a short form used in the following parts). They are either often present while driving or similar to some behaviors in the vehicle. Different modalities (auditory, visual, auditory+visual) are involved in these tasks. In the following, all these tasks will be described in details.

Music: In this task, participants just need to listen to some pieces of quiet music. This often happens while driving.

News: In this task, participants need to listen to some pieces of news. This is also a frequent task during driving.

Talk: In this task, participants are instructed to listen to two simultaneously presented voice recordings, a female news reader and a recording of two males talking. This mimicks a situation in which several passengers are talking whilst news is playing in the radio (similar to the task investigated by Kohlmorgen et al. (2007)). The participants were instructed to follow the latter, i.e. the two males' talk. To verify whether the participants were engaged or not, they had to answer related questions to the talk.

Movie: In the autonomous driving scenarios, watching movies in the vehicle will be usual. In this task, participants need to watch several movies and answer related question after each trial.

Letter: In this task, participants are instructed to read several little-known English fairy tales and count the occurrence of a certain letter. Here, they don't need to read and understand the text, but only have to identify the shape of the letters, which is relatively easy.

Adjective: In this task, participants are instructed to read several little-known English fairy tales and count the number of adjectives in the text. For this task, the text has to be understood in order to identify adjectives, which is rather demanding. We chose fairy tales since it's easier to keep up participant's motivation, and little-known English fairy tales could make sure that the participants don't know the text yet.

Calculation: In this task, participants need to silently count down in steps of b, starting from an initially given three-digit random number a. After a certain time period, the formula *a-b* disappeared and the participants were asked for the final result.

8.2.2 Secondary task

The secondary task is an auditory oddball task, which has been frequently used to examine signal detection ability and vigilance level. In this task, participants were instructed to distinguish target stimuli from numerous distractors. Targets are normalized pure tones of 1000Hz, at 75 dB, and distractors are normalized pure tones of 500Hz at 75 dB, both with duration of 1s. Target's probability was relatively low (25%), while distractors appear more often (75%).

The two tones were presented in a quasi-random order. Subjects were instructed to button press with hand in response to target and to ignore the lower tones, and the speed and accuracy of response were equally stressed.

8.3 Pre-study: Manipulating workload of different non-driving tasks and vigilance under different workload conditions

The pre-study is aimed at

- determining drivers' subjectively perceived mental workload levels when undertaking different non-driving tasks and

- investigating whether the drivers will miss more auditory alarms under higher workload in autonomous driving situations, i.e. whether the driver is less capable of readiness to take over under higher workload.

As we want to classify between different workload levels with EEG in the following study, it is important to set ground truth (labels) to the workload levels of different non-driving tasks, through assessing their subjectively reported mental workload levels with a standard workload questionnaire.

9 participants whose age ranged from 18 to 24 (M=21, S=2.24) participated in the pre-study. The experiment took place in a simulated real driving vehicle, with steering wheel, gas pedal, brake pedal and other control elements. The driving scene was projected onto a large screen around 2 meters in front of the driver. The NASA-TLX questionnaire (Chinese version from Xiao et al., 2005) and the one-itemed SEA-Scale (self-translated from the original German version into Chinese, see Appendix) on subjective load was used to investigate the perceived task load of each non-driving task.

During the experiment, the participants were firstly asked to read the instruction of this experiment, illustrating all different types of primary tasks and the secondary task. They had a few minutes to practice and get familiar with all primary tasks and the simultaneously presented oddball task. In the experiment process, participants were instructed to perform successively all seven different kinds of non-driving tasks, whose sequence was randomized. At the same time, they performed the secondary task (auditory oddball task) by pressing a certain button immediately after hearing the target stimulus among the distractor stimuli. Data on button press were recorded. After each session, participants reported their subjective workload by answering the NASA-TLX questionnaire and SEA-Scale.

Data analysis includes the NASA-TLX and SEA-Scale on subjective load and performance of button press (mainly the detection rate of target stimuli from the distractor stimuli).

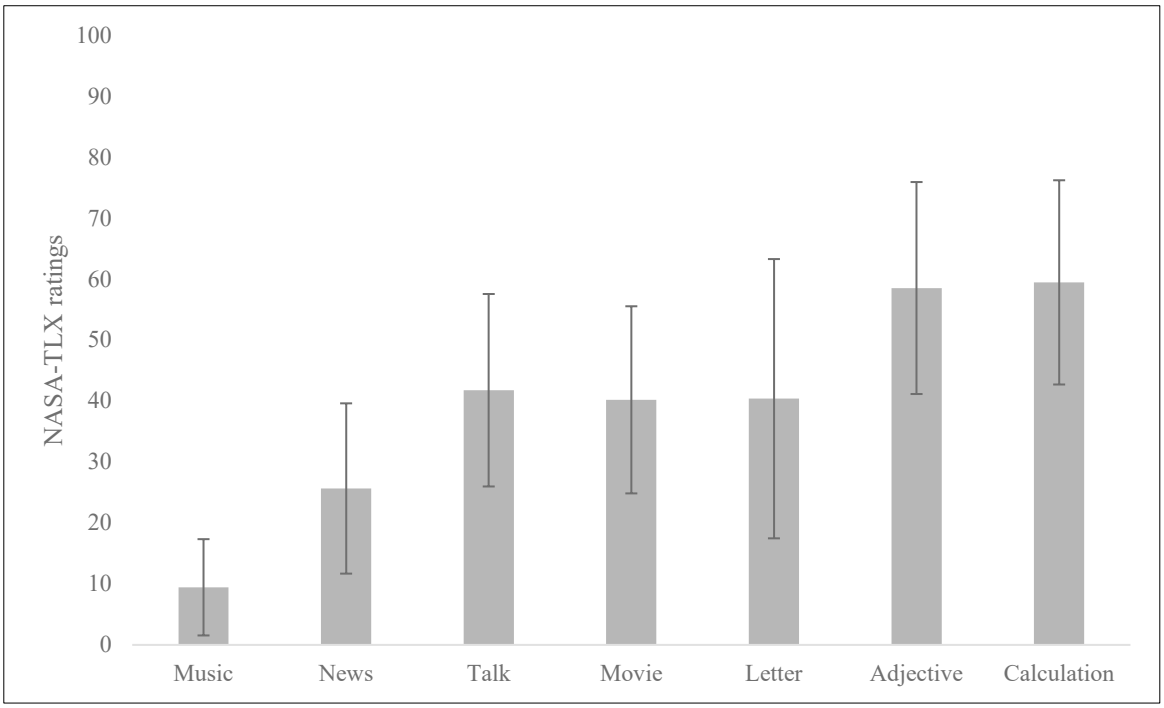


Figure 8.1 NASA-TLX ratings of different tasks

Table 8.1 p-values of t-test of NASA-TLX rating scores

Low – Medium		**Low – High**		**Medium – High**	
Music – Talk	$p < 0.001$	Music vs. Adjective	$p < 0.001$	Talk vs. Adjective	$p = 0.014$
Music –Movie	$p < 0.001$	Music vs. Calculation	$p < 0.001$	Talk vs. Calculation	$p = 0.009$
Music –Letter	$p = 0.001$	News vs. Adjective	$p = 0.001$	Movie vs. Adjective	$p = 0.003$
News –Talk	$p = 0.007$	News vs. Calculation	$p < 0.001$	Movie vs. Calculation	$p = 0.004$
News –Movie	$p = 0.013$			Letter vs. Adjective	$p = 0.045$
News –Letter	$p = 0.009$			Letter vs. Calculation	$p = 0.002$

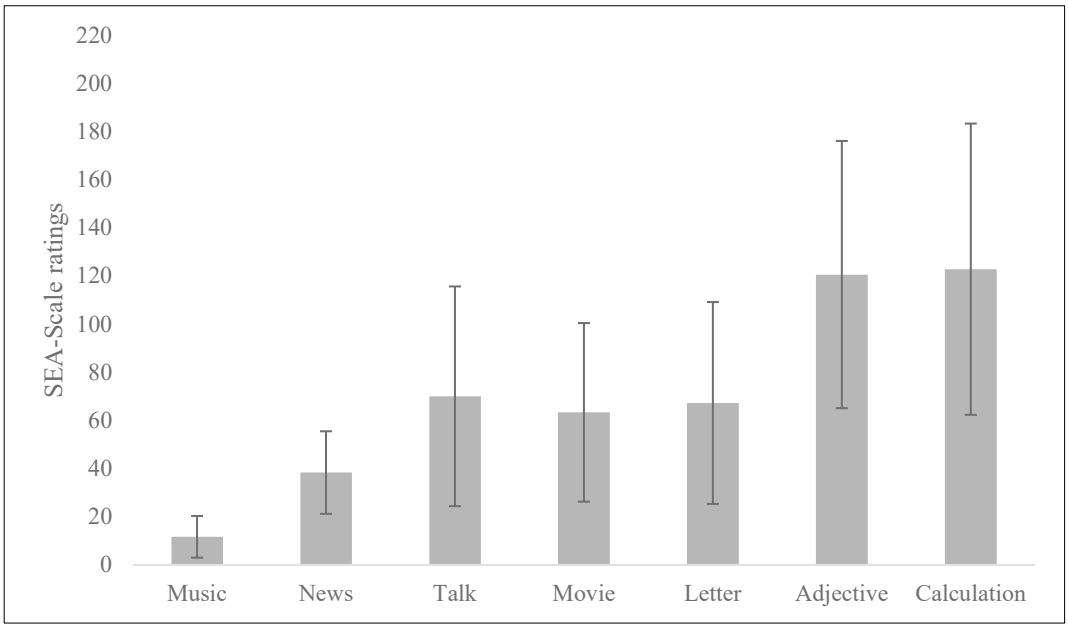


Figure 8.2 SEA-Scale ratings of different tasks

Table 8.2 p-values of t-test of SEA-Scale rating scores

Low – Medium		**Low – High**		**Medium – High**	
Music – Talk	p = 0.004	Music vs. Adjective	p < 0.001	Talk vs. Adjective	p = 0.019
Music –Movie	p = 0.001	Music vs. Calculation	p < 0.001	Talk vs. Calculation	p = 0.015
Music –Letter	p = 0.003	News vs. Adjective	p = 0.001	Movie vs. Adjective	p = 0.005
News –Talk	p = 0.037	News vs. Calculation	p = 0.004	Movie vs. Calculation	p = 0.026
News –Movie	p = 0.033			Letter vs. Adjective	p = 0.033
News –Letter	p = 0.047			Letter vs. Calculation	p = 0.009

According to the subjective reports (see Figure 8.1 & 8.2, Table 8.1 & 8.2), we could see differences in perceived mental workload among the 7 non-driving tasks. Results from NASA-TLX and SEA-Scale are rather consistent with each other. Thus, we divide these 7 tasks into 3 group: low workload tasks – music & news; medium workload tasks – talk, movie & letter; high workload tasks – adjective & calculation. For each task pairs in the low-medium, low-high and medium-high comparisons, t-test was performed and its

corresponding p-value was calculated. For all comparisons, there were significant differences between different task conditions, no matter in NASA-TLX or SEA-Scale ratings.

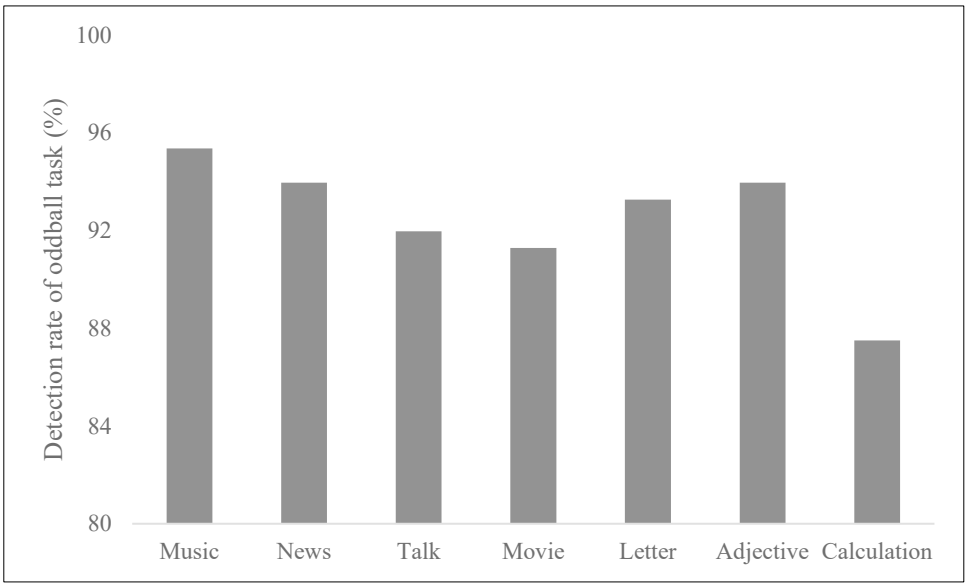


Figure 8.3 Detection rate in auditory oddball task

Regarding the detection rate in auditory oddball task (see Figure 8.3), we could also see that in the calculation task which induces higher workload, there were relative more misses and the detection rate of target stimuli was lower, compared with those in the lower workload tasks. This is consistent with our hypothesis and could provide support to following experiment. However, the detection rate in the adjective task is not significantly lower than other low workload tasks, or even a bit higher than some tasks with medium workload. A small number of samples might be the reason. Furthermore, because of this, we also lack statistical proof for the difference between calculation task and other tasks. Therefore, we'll take further steps to investigate these in the following experiment.

9. Experiment 2a – Detection of driver's mental workload with EEG

This experiment is to investigate whether it is possible to classify between different mental workload levels with EEG and apply the task-independent classifier (which has been proven in Experiment 1) in identifying whether the driver is under heavy or low workload across different non-driving activities.

9.1 Methods

9.1.1 Participants

In total, 15 participants (6 females and 9 males) between the ages of 18-24 (M=20.2, SD=1.57) took part in this experiment. Prerequisite for participation was the possession of a valid driving license and normal or corrected-to-normal vision. All participants were reported to be free of illness and medication. Caffeine, tobacco, and alcohol were prohibited prior to participating in the experiments. Participants were rewarded with a cash payment for their participation.

9.1.2 Experimental design

This experiment featured a one-factorial within-subject design. The *independent variable* was objective workload level, which was varied through task difficulties in different non-driving tasks. It consisted of three levels and they were validated by assessing the subjective workload level of these non-driving tasks in the Pre-study (see Chapter 8.3). The 'low' task difficulty condition has the *Music* and *News* tasks. These tasks are very easy to accomplish and don't demand much effort, and according to the subjective measurement on workload, they induce relatively low mental workload. The *Movie* and *Letter* tasks have 'medium' task difficulty. In processing these tasks, it requires more resources compared to the easy tasks and based on the subjective workload results in pre-study they also arouse medium mental workload. The 'high' task difficulty condition has the *Adjective* and *Calculation* tasks. Besides, as we want to have more 'high'-difficulty tasks, we increased the difficulty of *Talk* task by augmenting the volume of news reader in the background and thus making it more difficult for participants to follow the talk (in the pre-study the *Talk* task had medium task difficulty). The task difficulty of modified *Talk* task will also be verified in this experiment again via the subjective ratings on the task. All the three 'high'-difficulty tasks are very complex and have very high demand in mental workload.

The dependent variables in this experiment were subjective ratings on mental workload (scores of NASA-TLX and SEA-Scale), classification accuracy of direct training EEG data of non-driving tasks, performance of the task-independent workload classifier, as well as continuous classifier outputs of the task-independent classifier in the range between 0 to 1 as metric for mental workload. Subjective ratings also play the role of ground truth and promote the determination of labels of classes for the classification.

9.1.3 Procedure

Following a general introduction to the experimental procedure, participants were provided with an informed consent form and if no disagreement they signed this form. Following this, participants were asked to fill out a demographic questionnaire. After that, the EEG electrode cap was placed on the participant and conductive gel was applied and impedance of electrodes were adjusted for proper use. In order to reduce impedance and ensure better conduction performance, all female participants washed their hairs the night before the experiment and all males washed hairs shortly before the experiment. After all these preparation work finished, participants were seated in the driving simulator and began the real experiment phase.

The experiment involved two sessions: training session and testing session. In the training session, a calibration task was performed. Same as in Experiment 1, the calibration task has two conditions, the relaxing condition and mental calculation condition. The calibration task is around ten minutes long and its procedure is same as described in Experiment 1 (see Chapter 7.3). The testing session consists of seven blocks, and in each block a different non-driving task was performed, corresponding to different task difficulties. In each block, participants need to perform two tasks simultaneously. One is a kind of non-driving task for inducing workload, and the other is the auditory oddball task. A precise description of these tasks is given in Chapter 8.2. The sequence of non-driving tasks in these seven blocks was randomized. After each block, a questionnaire on mental workload was filled out, which consisted of both NASA-TLX (Chinese version) and one-itemed SEA-Scale (self-translated Chinese version). Until all the seven blocks in the testing session were completed, the experiment phase was over and participants could remove the electrode cap and wash their hair. The electrode cap was also cleaned.

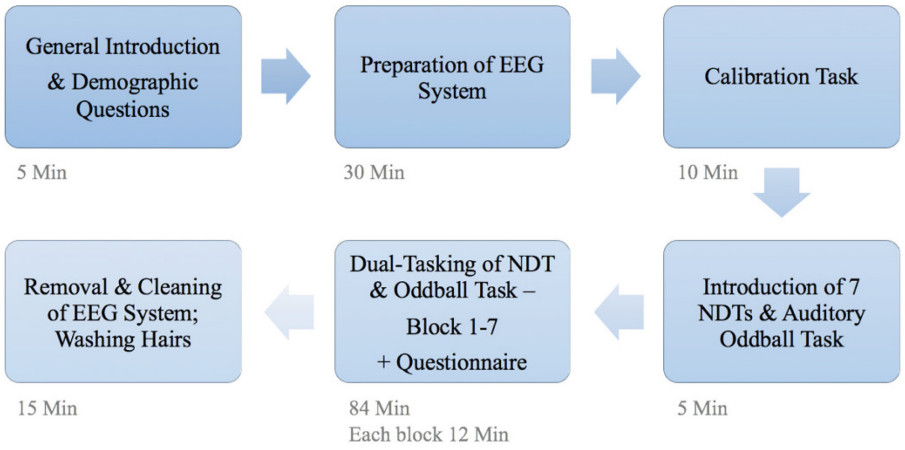


Figure 9.1 General procedure of Experiment 2

9.1.4 Experiment apparatus

This experiment was conducted in a static driving simulator, which is a real Volkswagen vehicle, with steering wheel, gas/brake pedals and other control elements. The simulated autonomous driving environment was developed by BCMI team from Shanghai Jiaotong University[10] and projected onto a large screen around 2 meters in front of the driver. In the autonomous driving situation, the vehicle was driving on a highway at the constant speed of 80 km/h, which was set up beforehand.

[10] Homepage of BCMI team: http://bcmi.sjtu.edu.cn/

Figure 9.2 Driving simulator – A real driving vehicle

Brain activity was recorded with a 62 active Ag/AgCl electrode system by ESI NeuroScan. Electrodes on the cap are mounted according to the higher-resolution international 10-20 system. EEG signals were sampled at 1000 Hz and wide-band filtered (0.5-70 Hz).

Figure 9.3 NeuroScan wet electrode cap

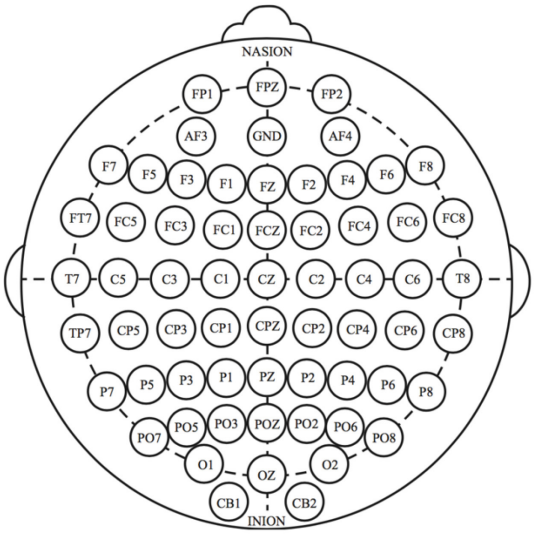


Figure 9.4 Higher-resolution international 10-20 system

The non-driving tasks were programmed in Python and displayed with the software PsychoPy (version 3.0) on a 17-inch PC screen at the position of Infotainment in vehicle. Music, news and talks were played with a loud speaker. A small number keyboard was placed onto the steering wheel and one of the buttons was set as reaction for oddball task. Markers for the ‘low’, ‘medium’ and ‘high’ conditions were transmitted in real time via parallel port to the the Scan software (version 4.5) and synchronized with the recorded EEG signal.

Figure 9.5 A participant wearing EEG system seated in the simulator vehicle

9.1.5 Data analysis

The methods used here are similar to those in Experiment 1 (see Chapter 8.4.5). Firstly, subjective ratings on NASA-TLX and SEA-Scale were computed over all participants for each task session. Pairwise t-tests of all difficulty pairs were statistically analyzed. After the differences in three task difficulty conditions were verified, pairwise classifications between each two of three conditions (low vs. high, low vs. medium, medium vs. high) were then performed directly on data of tasks in these three workload conditions. After that, same as in Experiment 1, individual classifiers were also trained on a standardized calibration task and applied on data of all non-driving tasks and to classify between different workload conditions. During the process of classification, all EEG data were divided in consecutive 1-second epochs of different conditions (low/medium/high) data. The filter bank common spatial patterns (FBCSP) was used to extract features in the frequency bands of theta (4-7 Hz) and alpha (8-13 Hz) with three patterns per band. To differentiate the classes, linear discriminant analysis (LDA) was applied with a 5-fold nested cross-validation with margins of 5. Based on these, classification accuracies of different applications were computed. Moreover, classifier outputs of different non-driving tasks were calculated, and the differences between conditions were compared with permutation tests.

Calibration and classification were done using the open-source MATLAB-based toolbox BCILAB (version 1.2). Statistical analysis was performed with SPSS Statistics (version 26.0.0.0).

9.2 Results

9.2.1 Subjective measures

Measures on subjective mental workload are performed as method to ascertain ground truth of task difficulties and give the seven non-driving tasks corresponding labels in the classification process. Based on pre-study and results on subjective rating score of NASA-TLX and SEA-Scale questionnaires (see Table 8.1 & 8.2, Figure 8.1 & 8.2), we could see variations among all seven non-driving tasks. Subjective ratings increased as a function of task difficulty and the seven tasks can be correspondingly divided into three groups: low, medium, and high. In the Music and News tasks, participants perceived relative low workload; in the Movie and Letter tasks, they had medium workload; in the Talk, Adjective and Calculation tasks, there were relative high workload on them.

Table 9.1 Subjective reporting scores of NASA-TLX & SEA-Scale

		Music	**News**	**Movie**	**Letter**	**Talk**	**Adjective**	**Calculation**
NASA-TLX	M	14.89	21.28	25.17	29.06	50.56	43.11	52.67
	SD	8.37	11.88	11.95	10.66	13.13	13.24	13.60
SEA-Scale	M	15.43	40.60	46.05	47.19	112.97	97.83	120.97
	SD	11.89	34.30	32.09	30.92	51.15	50.86	47.39

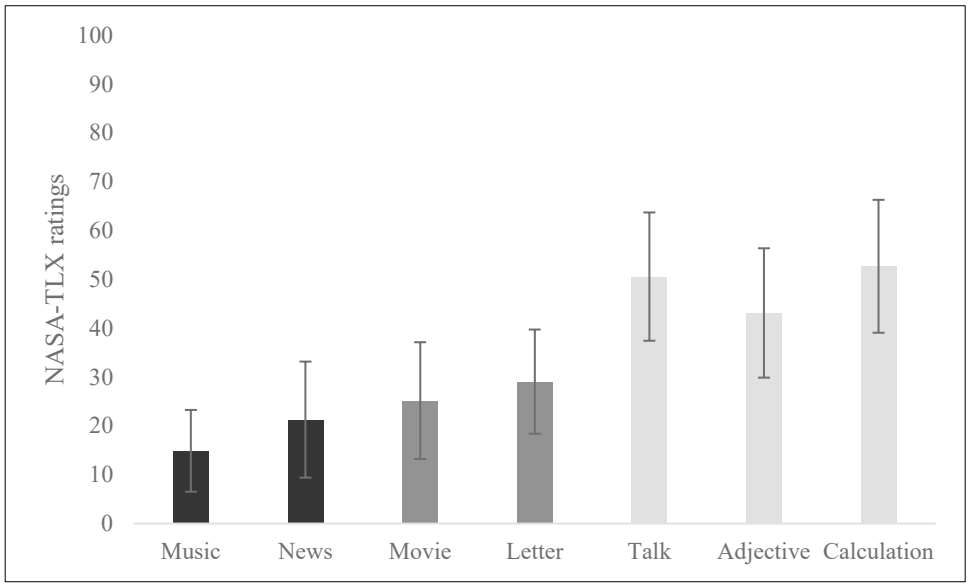


Figure 9.6 Subjective ratings of NASA-TLX

As seen from Table 9.1 and Figure 9.6, although NASA-TLX ratings of Talk, Adjective and Calculation tasks are already higher than other tasks, their average ratings are all around 50, which is just at medium level along the score range. However, original NASA-TLX was mostly applied in aviation area (Hart & Staveland, 1988) and tasks involved in the aviation area are much more demanding and complex. In comparison to tasks in aviation, cognitive tasks in laboratory are always more simple, therefore, ratings of high workload tasks around 50 are relatively reasonable.

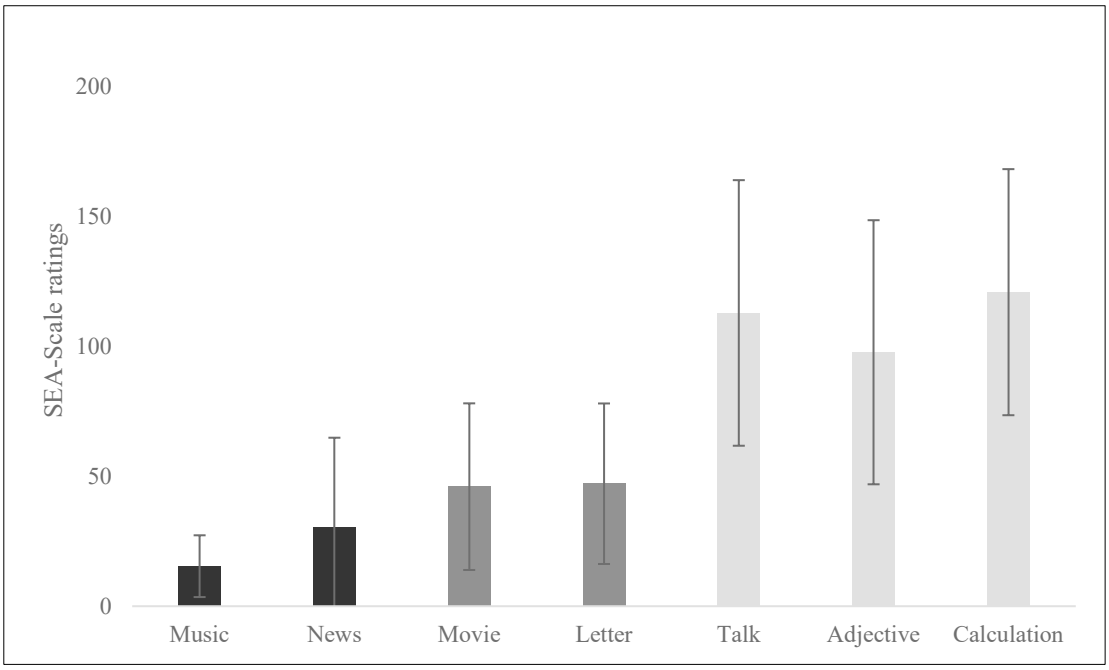


Figure 9.7 Subjective ratings of SEA-Scale

For the NASA-TLX and SEA-Scale rating scores in all non-driving tasks, pairwise t-tests were performed for all comparisons between low vs. medium condition, low vs. high condition, and medium vs. high conditions and their p-values were calculated. From Table 9.2 we could see, that differences of NASA-TLX rating scores in all comparing pairs were significant. For SEA-Scale ratings (see Table 9.3), besides the comparisons of news vs. movie tasks and news vs. letter tasks, there were significant differences in all other task pairs.

Table 9.2 p-values of t-tests of NASA-TLX rating scores

Low – Medium		**Low – High**		**Medium – High**	
Music –Movie	$p < 0.001$	Music vs. Talk	$p < 0.001$	Movie vs. Talk	$p = 0.019$
Music –Letter	$p = 0.002$	Music vs. Adjective	$p < 0.001$	Movie vs. Adjective	$p = 0.025$
News –Movie	$p = 0.032$	Music vs. Calculation	$p < 0.001$	Movie vs. Calculation	$p < 0.001$
News –Letter	$p = 0.049$	News vs. Talk	$p < 0.001$	Letter vs. Talk	$p < 0.001$
		News vs. Adjective	$p < 0.001$	Letter vs. Adjective	$p = 0.001$
		News vs. Calculation	$p < 0.001$	Letter vs. Calculation	$p < 0.001$

Table 9.3 p-values of t-tests of SEA scale ratings

Low – Medium		**Low – High**		**Medium – High**	
Music –Movie	$p < 0.001$	Music vs. Talk	$p < 0.001$	Movie vs. Talk	$p = 0.001$
Music –Letter	$p = 0.001$	Music vs. Adjective	$p < 0.001$	Movie vs. Adjective	$p = 0.014$
News –Movie	$p = 0.143$	Music vs. Calculation	$p < 0.001$	Movie vs. Calculation	$p = 0.003$
News –Letter	$p = 0.582$	News vs. Talk	$p < 0.001$	Letter vs. Talk	$p < 0.001$
		News vs. Adjective	$p = 0.005$	Letter vs. Adjective	$p = 0.001$
		News vs. Calculation	$p < 0.001$	Letter vs. Calculation	$p < 0.001$

(p-value highlighted with grey means insignificant)

9.2.2 Classification accuracy of direct training on non-driving tasks

Table 9.4 list the classification accuracies of pairwise Low vs. High conditions, which means, brain signals from low and high difficulty task pairs were directly trained and classified (low difficulty tasks: Music and News tasks; high difficulty tasks: Talk, Adjective and Calculation tasks). The mean estimated classification accuracy over participants of these task pairs ranged from 70.4% to 89.3% (significance is reached at 55% or above). Overall, classification accuracies between Low and High condition tasks are acceptable.

Table 9.4 Classification accuracies of Low vs. High condition tasks

Participant	Music vs. Talk	Music vs. Adjective	Music vs. Calculation	News vs. Talk	News vs. Adjective	News vs. Calculation
1	60.47	94.58	80.55	61.54	89.75	87.16
2	67.90	95.17	83.69	57.80	83.04	79.00
3	65.72	93.34	91.44	63.98	97.76	95.18
4	73.85	92.44	91.93	70.41	92.33	91.29
5	72.11	79.53	89.29	64.55	88.12	87.04
6	89.31	92.29	97.29	73.01	85.06	95.10
7	68.34	89.54	89.41	76.67	74.85	57.88
8	98.90	81.66	91.64	76.10	99.21	90.35
9	82.80	94.03	81.25	72.76	71.33	88.88
10	97.00	99.67	73.78	74.15	98.90	91.12
11	48.56	56.47	98.39	97.15	99.15	78.72
12	74.63	96.06	93.65	56.75	97.36	98.54
13	66.25	78.79	78.02	73.58	82.29	79.14
14	94.65	98.65	98.11	85.37	95.00	99.83
15	48.97	90.73	86.69	51.87	84.70	81.15
M	73.97	88.86	88.34	70.38	89.26	86.69
SD	15.92	11.03	7.53	11.62	8.97	10.54

Table 9.5 list the classification accuracies of pairwise Low vs. Medium conditions (low difficulty tasks: Music and News tasks; medium difficulty tasks: Letter and Movie tasks). Due to some technical problem when performing the Movie task, some participants (participant 1, 3, 4, 5, 10, 11) didn't complete the Movie task and only data of 9 participants were recorded. We could see from this table that all the four low and medium difficulty task pairs reached mean estimated classification accuracy above 80%, ranging from 82.6% to 87.6%.

Table 9.5 Classification accuracies of Low vs. Medium conditions

Participant	**Music vs. Letter**	**News vs. Letter**	**Music vs. Movie**	**News vs. Movie**
1	84.16	86.78		
2	94.95	85.49	97.72	76.57
3	88.77	91.16		
4	96.63	90.44		
5	83.68	82.30		
6	89.71	73.85	98.29	97.54
7	76.38	70.54	76.36	78.05
8	97.30	93.20	77.63	99.68
9	84.55	82.34	88.11	72.53
10	89.33	91.22		
11	80.77	99.69		
12	91.27	97.59	80.05	91.58
13	86.04	93.37	69.40	70.00
14	94.75	97.51	81.41	99.05
15	76.29	77.61	78.52	58.00
M	87.64	87.54	83.05	82.56
SD	6.77	8.80	9.78	14.95

Table 9.6 list the classification accuracies of pairwise Medium vs. High conditions (medium difficulty tasks: Letter task; high difficulty task: Talk, Adjective and Calculation). As the data of Movie task were not complete, we only considered the Letter task in the medium condition here. The mean estimated classification accuracy over participants of these task pairs ranged from 75.3% to 91.4%.

Table 9.6 Classification accuracies of Medium vs. High conditions

Participant	**Letter vs. Talk**	**Letter vs. Adjective**	**Letter vs. Calculation**
1	87.10	88.44	89.30
2	90.70	69.60	90.22
3	92.50	93.56	89.42
4	93.38	73.02	97.33
5	78.25	57.66	91.63
6	72.41	65.48	89.58
7	78.14	62.67	69.39
8	82.94	99.10	98.90
9	79.06	52.41	89.74
10	90.98	99.70	95.84
11	79.08	69.65	99.56
12	97.02	75.10	96.58
13	93.12	63.36	96.65
14	98.66	82.99	99.79
15	78.69	76.54	76.50
M	86.14	75.29	91.36
SD	8.20	14.72	8.56

9.2.3 Performance of task-independent workload classifier

Table 9.7 lists the classification accuracies of a classifier trained on data from the calibration task, and applied to *the data of low vs. high difficulty tasks*. The mean estimated classification accuracy over participants in the calibration task reaches 76.7%±8.6. Applied to the testing tasks, however, the mean classification accuracies of different task pairs ranged from 57.8% to 62.3%.

Table 9.7 Classification accuracy after applying task-independent classifier (%)

Participant	Calibration	Music vs. Talk	Music vs. Adj	Music vs. Calc	News vs. Talk	News vs. Adj	News vs. Calc
1	89.5	63.0	71.9	76.3	57.1	73.3	60.5
2	78.3	57.0	62.7	72.3	56.1	69.3	72.3
3	84.0	62.8	68.3	68.0	64.6	65.0	57.5
4	79.8	57.8	64.9	71.6	57.5	68.6	51.7
5	88.3	65.1	75.6	74.8	58.4	71.8	50.4
6	71.3	52.8	48.1	45.7	48.8	52.7	48.6
7	60.8	57.7	48.7	42.7	65.0	59.7	55.5
8	84.7	50.1	53.5	61.1	55.1	58.1	65.6
9	78.0	53.6	43.6	64.3	51.0	61.3	79.4
10	68.5	50.7	52.3	47.3	55.8	54.3	72.3
11	67.8	57.7	69.3	63.4	54.3	60.4	49.1
12	83.4	60.2	72.9	62.8	63.7	59.8	60.1
13	76.4	63.1	58.1	51.2	62.0	48.2	59.8
14	74.5	68.6	63.4	59.0	57.8	56.0	75.8
15	65.6	58.5	55.0	56.1	59.9	53.1	75.3
M	76.7	58.6	60.5	61.1	57.8	60.8	62.3
SD	8.6	5.3	10.1	10.7	4.7	7.5	10.5

Table 9.8 lists the mean classifier output values produced when applying the calibrated classifier to the data of all testing tasks. Values could vary between 0 and 1, and represent the classifier's predictions, with 0.5 representing complete uncertainty between the classes, 0 representing certainty for class 1 (i.e. the classifier predicts a 100% chance of the data belonging to class 1, low workload) and 1 certainty for class 2 (high workload).

Table 9.8 Averaged classifier outputs of each task across participants

	Music	**News**	**Talk**	**Movie**	**Letter**	**Adjective**	**Calculation**
M	0.49	0.51	0.64	0.59	0.64	0.70	0.69
SD	0.12	0.10	0.11	0.15	0.15	0.14	0.14

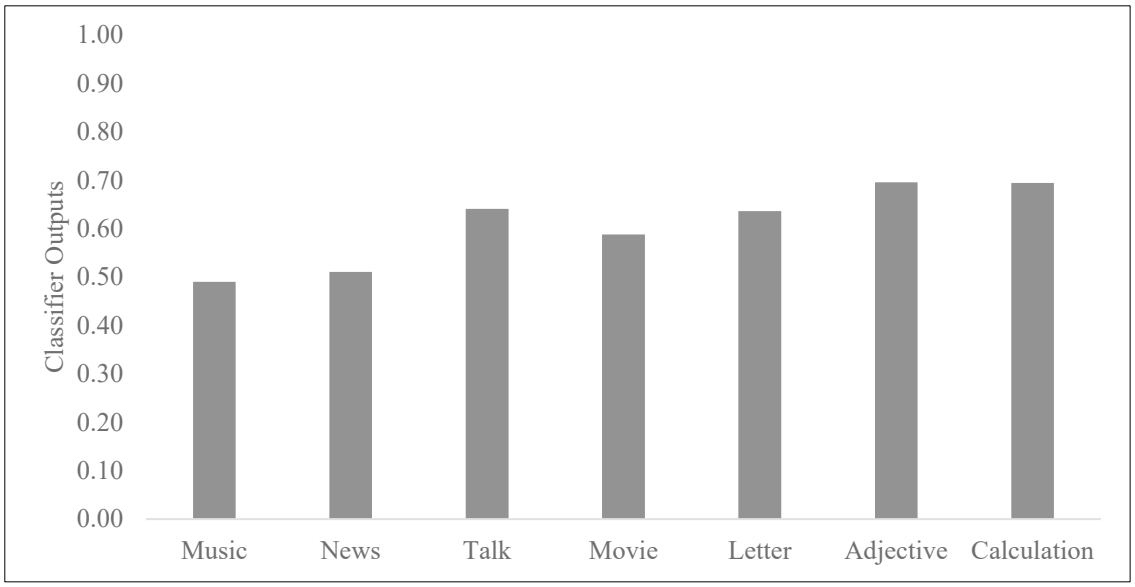


Figure 9.8 Averaged classifier outputs of each task

Figure 9.8 presents the mean classifier output across all participants. We could see that the mean output for Music task is below the middle line (0.5), whereas the mean outputs of other six tasks are all above 0.5.

Table 9.9 lists the results of permutation tests comparing the output within each participant between high and low workload conditions. For each task pair, over 73.3% of the participants had significantly different classifier outputs between conditions. Over all participants, in 83.3% of all tasks the differences between high and low conditions were significant.

Table 9.9 Results of permutation tests on outputs of High vs. Low workload conditions

Participants	**P-value of Permutation Test**						**Percentage of Task Pairs w. Significance (%)**
	Music vs. Talk	**Music vs. Adj**	**Music vs. Calc**	**News vs. Talk**	**News vs. Adj**	**News vs. Calc**	
1	p=0.007**	p < 0.001***	p < 0.001***	p=0.170	p < 0.001***	p < 0.001***	83.3
2	p=0.038*	p < 0.001***	p < 0.001***	p=0.002*	p < 0.001***	p < 0.001***	100
3	p < 0.001***	p < 0.001***	p < 0.001***	p < 0.001***	p < 0.001***	p=0.474	83.3
4	p=0.048*	p < 0.001***	p < 0.001***	p=0.342	p < 0.001***	p=0.048*	83.3
5	p < 0.001***	p < 0.001***	p < 0.001***	p=0.326	p < 0.001***	p=0.099	66.7
6	p < 0.001***	p < 0.001***	p < 0.001***	p < 0.001***	p < 0.001***	p < 0.001***	100
7	p=0.507	p=0.235	p < 0.001***	p < 0.001***	p < 0.001***	p < 0.001***	66.7
8	p < 0.001***	p < 0.001***	p < 0.001***	p=0.001***	p < 0.001***	p < 0.001***	100
9	p < 0.001***	p < 0.001***	p < 0.001***	p < 0.001***	p < 0.001***	p < 0.001***	100
10	p < 0.001***	p=0.054	p < 0.001***	p=0.022*	p < 0.001***	p < 0.001***	83.3
11	p=0.002**	p < 0.001***	p < 0.001***	p < 0.001***	p < 0.001***	p < 0.001***	100
12	p=0.014*	p < 0.001***	p < 0.001***	p < 0.001***	p < 0.001***	p < 0.001***	100
13	p=0.598	p < 0.001***	p < 0.001***	p < 0.001***	p < 0.001***	p < 0.001***	83.3
14	p < 0.001***	p < 0.001***	0.010**	p=0.833	p=0.010**	p < 0.001***	83.3
15	p=0.505	p=0.692	p=0.633	p=0.050*	p=0.630	p < 0.001***	33.3
Percent. of particip. w. significance (%)	73.3	80	93.3	73.3	93.3	86.7	Average: 83.3

(p-value highlighted with grey means insignificant)

9.3 Discussion

The aim of this experiment was to validate the possibility of classifying different workload levels with brain signals. First of all, impact of task difficulty on subjective measures and the possibility of grouping the non-driving tasks into different classes for the further classification based on subjective measures will be discussed. Following that, data on direct training on different task pairs representing low-high, low-medium, and medium-high conditions will be interpreted. Afterwards, results on applying the task-independent workload classifier on testing tasks and its capability to differentiate between high and low conditions will be discussed.

9.3.1 Subjective measures

Subjective ratings from NASA-TLX and SEA-Scale confirmed the assumption that there were three different workload levels induced by these seven non-driving tasks. Consistent with the hypothesis, the overall workload ratings increased with increasing task difficulty and can be distinguished between all three difficulty levels.

Due to the fact that task difficulty is a major influencer of mental workload and based on results on subjective workload in relation to task difficulty, it can be assumed that these three conditions do impose three different workload levels, and can be further used as class labels for classification of EEG data.

9.3.2 Direct training classification accuracy

The classification accuracies between low-high, low-medium, and medium-high workload tasks are all acceptable. This is consistent with our assumption, that with the implemented methods, variations of mental workload can be indicated in EEG data. Changes in mental workload do elicit significant differences in the power of theta and alpha bands. As stated in Chapter 5, a number of studies (e.g. Brookings et al., 1996; Gevins et al., 1998; Smith et al., 2001) have demonstrated that human brain activity reflects an increase of mental workload by an increase in frontal theta power and a decrease in parietal alpha power. This have also been confirmed with illustrations of spatial patterns in Experiment 1 (see Chapter 8) when applying Filter Bank Common Spatial Patterns to extract features associated with mental workload. Good results on classification accuracies in directly training EEG data from different tasks in this experiment further prove the differentiability of brain signals and feasibility of the applied methods for workload detection.

Moreover, former studies using FBCSP and LDA for workload classification were more involved with cognitive tasks like n-back task and arithmetic tasks. In this experiment, the non-driving tasks were more

realistic and practical tasks, which are more likely to happen during driving or in the vehicle. Application of these methods in a more practical scenario were verified to be viable, and this extends the possibility of applying FBCSP and LDA in mental workload detection.

9.3.3 Task-independent workload classifier

As demonstrated in Experiment 1, it is possible to calibrate a general classifier on a standardized workload-inducing task, and apply this classifier to other testing tasks. We applied this method again in this experiment, with more practical driving-related tasks.

Although a direct transfer from the standardized calibration task to the practical testing tasks inducing workload is not able to distinguish between low and high workload levels acceptably, we could see obvious increase in classifier outputs when observing from low workload tasks to medium and to high workload tasks. Apart from outputs of Music task, averaged classifier outputs of all the other non-driving tasks are above 0.5. As stated in Experiment 1 that the default separating line of high and low workload conditions is 0.5, a direct classification, which counts simply the percentage of trials below or above this default separating threshold of 0.5 for each class, is unreliable and not suitable here. Whereas, results of permutation tests on different task pairs of high and low workload conditions show that, for most task pairs and most participants (83.3%) there were significant differences between high and low conditions. This again confirms our statement in Chapter 8.

Furthermore, it is proven that calibrating a classifier on a continuous subtraction task and applying it to other practical non-driving tasks which induce different workload levels, is sufficient to indicate the mental workload levels of these tasks with a continuous metric of classifier outputs, which are differentiable between different task conditions.

In previous studies, to detect mental workload in real time it is of necessity to firstly train a classifier on the task and then apply this classifier on the same task, i.e. a classifier for each testing task need to be calibrated. This process is very time-consuming and cumbersome. This task-independent classifier is aimed at simplifying the process of online workload detection, by generating a universal, task-independent workload classifier for detection of mental workload of different testing tasks. The tasks investigated in Experiment 1 were all tasks conducted in a well-controlled laboratory. What more important in this experiment is that it is possible to apply this task-independent workload classifier in more realistic tasks and in real-world applications. Results from this experiment strongly support this possibility of real-life applications.

9.4 Summary

Overall, all dependent variables are sensitive to differences in task difficulty. Subjective ratings in both NASA-TLX and SEA-Scale can significantly distinguish between low-medium, low-high, and medium-high conditions. Classification accuracies of direct training on tasks of different conditions confirmed the differentiability of workload levels, and also proved the feasibility of implemented methods (FBCSP & LDA) for brain signal analysis. Furthermore, the application of a task-independent workload classifier was able to indicate mental workload with classifier outputs, which can significantly differentiate between classes. This task-independent workload classifier was also applicable in more realistic tasks and in more practical scenarios.

Although the task-independent classifier greatly simplifies the calibration process and optimize the application of online workload detection, the utilization of a gel-based EEG system is inconvenient due to long preparation time and limited portability. A portable dry electrode system with less electrodes would be a better solution. However, since such dry electrode systems are very sensitive to artifacts and have poorer data quality, its feasibility still need to be investigated in the future.

10. Experiment 2b – Vigilance detection with EEG

In many critical situations, operators may be unable to detect alarms due to overload and thus induced decreased vigilance. As stated by many researchers (e.g. Helton & Russell, 2012), engaging in a task leads to a depletion of mental resources, and consequently inducing a decrease in vigilance. The operator may probably neglect warnings especially under high workload settings, and the unattended displayed signals may fail to reach human operator's awareness. Moreover, different workload levels also have different impact on the ability to detect signals.

Therefore, in this experiment, we investigated the detection of vigilance with EEG in the highly automated driving situation. Workload's influence on vigilance level, i.e. performance of the auditory oddball task, was also examined.

To conclude, this experiment is aimed at

- investigating whether it is possible to detect signal perceptions among distractors based on EEG analysis;
- validating the influence of mental workload on vigilance level, which means, under higher mental workload condition less target signals can be detected from distractors, and vice verse.

10.1 Methods

As this experiment is an extension to Experiment 2a (Chapter 9) and was carried out together with tasks in Experiment 2a at the same time, most of the settings in this experiment are same as in Experiment 2a. Hence, in this chapter the same parts will be skipped and we'll focus on the different parts.

10.1.1 Participants

The participants are the same as in Experiment 2a (Chapter 9.1.1) and their general information is identical as well.

10.1.2 Procedure

The main part of this experiment is the auditory oddball task, in which participants were instructed to distinguish target stimuli from distractors. Since this experiment was embedded in Experiment 2a, the

oddball task was included in all seven non-driving task sessions. That is to say, in each session, there were dual tasks: the primary task is one of the non-driving tasks, and the secondary task is the auditory oddball task (as illustrated in Figure 10.1). Participants were required to perform the secondary task under the condition of guaranteeing the performance of the primary task.

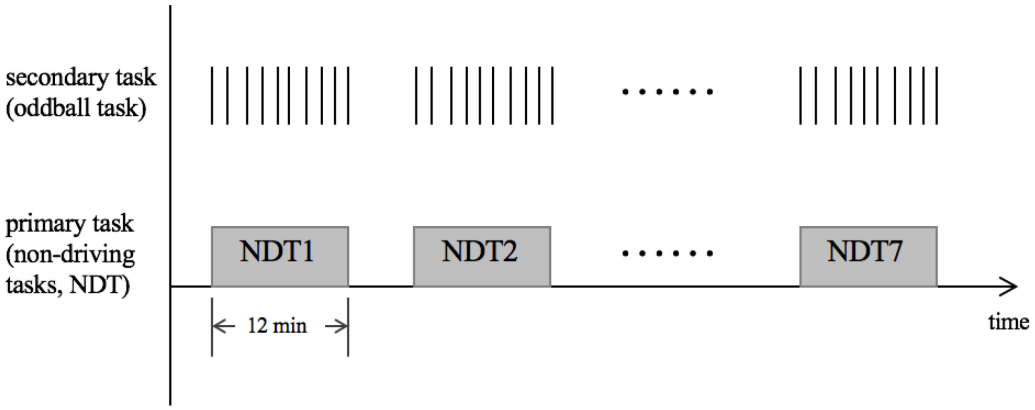


Figure 10.1 Illustration of procedure of dual tasks

In the single oddball task, there were two types of stimuli: targets and distractors. The type of auditory stimuli was varied through the pitch, i.e. the frequency of displayed auditory tones. Targets are normalized pure tones of 1000Hz, at 75 dB, and distractors are normalized pure tones of 500Hz at 75 dB (see also Williams et al., 2005), both with duration of 1s. Target probability was relatively low (25%), while distractors appear more often (75%). The two types of tones were presented in a quasi-random order. In between two tones, there was a random inter-stimulus interval (ISI) varying from 5 to 10 seconds, with an average of 7.5s. Taken the appearance rate of targets into consideration, a target was present about every 30 seconds. This interval is determined on account of German safety driving circuit (in German: Sicherheitsfahrschaltung, for short Sifa) regulation for trains in DIN 0119-207-5 (2016). In this regulation, the time interval between two safety driving inquiries should be approximately 30 seconds (Schmidt, 2018; Kohlmorgen, 2007). Participants were instructed to press a certain button with hand in response to targets while distractors should be ignored. During the oddball task, the speed and accuracy of response were equally stressed.

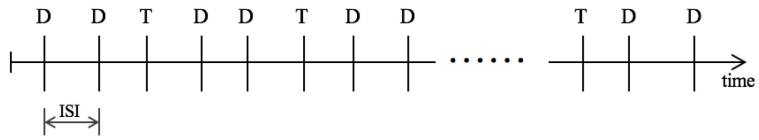


Figure 10.2 Procedure of the oddball task in one dual task block. The sequence of targets and distractors in this figure is merely an example. The inter-stimulus interval (ISI) was randomly chosen from 5 to 10 seconds, so that participants had to react every 30 seconds on average.

10.1.3 Data acquisition

Same as Experiment 2a, this experiment was also conducted in a static driving simulator. The auditory oddball task was displayed via a stereo inside of the vehicle, and as reaction to the targets, participants were instructed to press a button on a keypad, which was fixed on the steering wheel. Reaction time and response accuracy of button press in the oddball task were collected. Meanwhile, brain signals were recorded continuously with a ESI NeuroScan system from 62 active Ag/AgCl scalp electrodes located according to the higher-resolution international 10-20 system, at a 1000 Hz sampling rate and with a 0.5-70 Hz band-pass filter.

10.1.4 Data analysis

Based on the responses relative to stimulus types, we can divide the stimulus into four groups: hits, misses, false alarms and correct rejections. Combining with stimulus types (targets & distractors) and different workload conditions of primary tasks, several pairs of stimulus classes were grouped:

- targets vs. distractors
- hits vs. misses vs. correct rejections (false alarms are not considered here due to very low rate)

For each pair of stimuli, a few analyses were performed as follows.

The data were down-sampled to 200 Hz, and filtered with a band-pass filter of 0.1-25 Hz. Data were then segmented into epochs of 1200ms long, starting 200ms before the onset of each stimulus. ERPs were computed and averaged ERP curves of different stimulus classes were depicted.

Moreover, in order to assess the difference of ERPs relative to stimulus response, a classification was carried out between hits and correct rejections. For the classification, the data were filtered with a band-pass filter of 0.1-15 Hz, and epochs starting 200ms before and ending 800ms after the onset of each stimulus were extracted. Then the ERP intervals of 300-650ms after stimulus presentation were used for classification. The characteristics were extracted by the "Windowed Means" paradigm, which divides the extracted epochs into consecutive time windows of 50ms length and calculates the average of the amplitudes within the time window (Blankertz et al., 2011). These characteristics were classified using linear discriminant analysis (LDA) with a 5-fold cross-validation with margins of 5. Classification accuracies were computed.

Besides, rates of hits, misses, false alarms and correct rejections were calculated. Miss rates in different task conditions were compared with t-tests and p-values were computed to indicate the significance. Average reaction times of hits and false alarms under all task conditions were also calculated.

ERPs were analyzed with the open-source MATLAB-based toolbox EEGLAB (version 13.6.5b), and classification was done using the open-source MATLAB-based toolbox BCILAB (version 1.2). Statistical analysis was performed with SPSS Statistics (version 26.0.0.0).

10.2 Results

10.2.1 ERP analysis and classification accuracy

Target vs. distractor

Figure 10.3 shows the grand average of ERP curves of targets and distractors at channel Fz, FCz, Cz and Pz from -200ms to 950ms relative to the stimuli presence, across all participants and all task trials. Statistical analysis results were showed in bold black lines in the x-axis of time. From around 500ms on, there were significant potential differences between target and distractors ($p<0.05$), and the significant differences last until around 800ms.

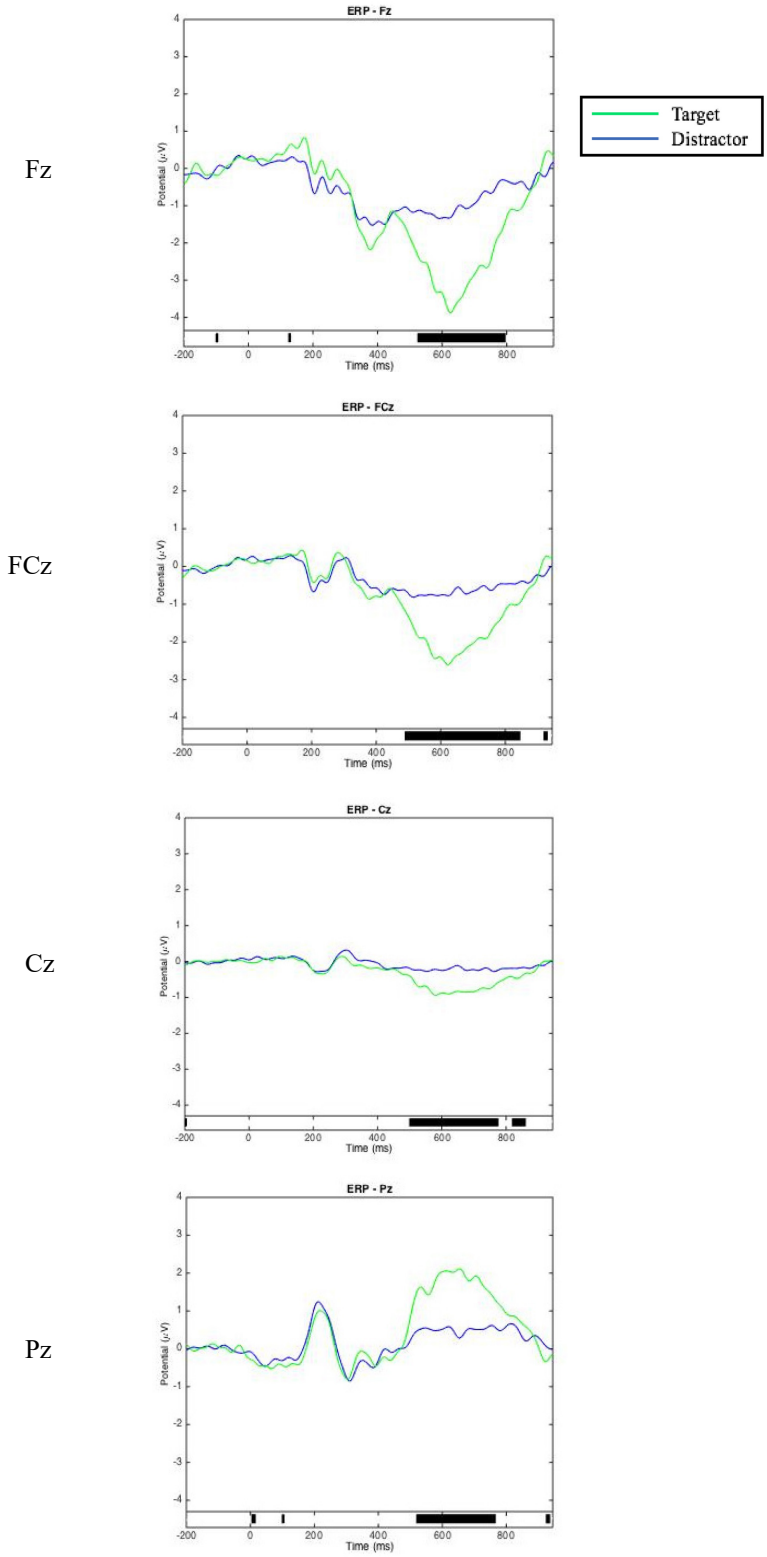


Figure 10.3 Grand average of targets and distractors at electrode positions Fz, FCz, Cz and Pz across all participants and all tasks

Table 10.1 Mean classification accuracies of targets vs. distractors across all participants in different task blocks (%)

	Music	**News**	**Talk**	**Movie**	**Letter**	**Adjective**	**Calculation**	**Overall**
M	64.10	63.08	65.86	64.80	66.13	69.67	66.13	65.68
SD	6.65	6.10	8.03	5.64	11.11	10.75	10.55	8.73

Table 10.1 lists the mean classification accuracies across all participants in different task blocks, when classifying brain signals from 350ms to 600ms of targets and distractors. The overall classification accuracy reaches 65.68%$\pm$8.73. Among them, the classification accuracy in the Adjective task block was the highest, in comparison to other task blocks. However, the difference of classification accuracy between different tasks are not obvious, and this makes sense since the classification of targets and distractors should not vary across different task blocks.

Hits vs. misses vs. correct rejections

Figure 10.4 shows the grand average of ERP curves of hits, misses and correct rejections at channel Fz, FCz, Cz and Pz from -200ms to 950ms relative to the stimuli presence, across all participants and all task trials. Between hits and other conditions, there were significant differences ($p<0.05$) in potential from 600ms on, at all these four electrodes. At FCz, Cz and Pz, there were significant differences around 200ms and at electrode Pz also significant around 400ms.

Looking at the ERP curve of hits at electrode Fz, there appeared a remarkable negative wave (NW) from 450 to 900ms after stimulus in the prefrontal electrode positions, and reached its peak at 650ms at electrode Fz. Furthermore, the ERP of hits showed a distinct P3 component at the parietal electrode Pz, and its maximal potential appeared about 600ms after the stimuli. This positive wave in hits shows significant difference ($p<0.05$) in the parietal positions in comparison with other conditions.

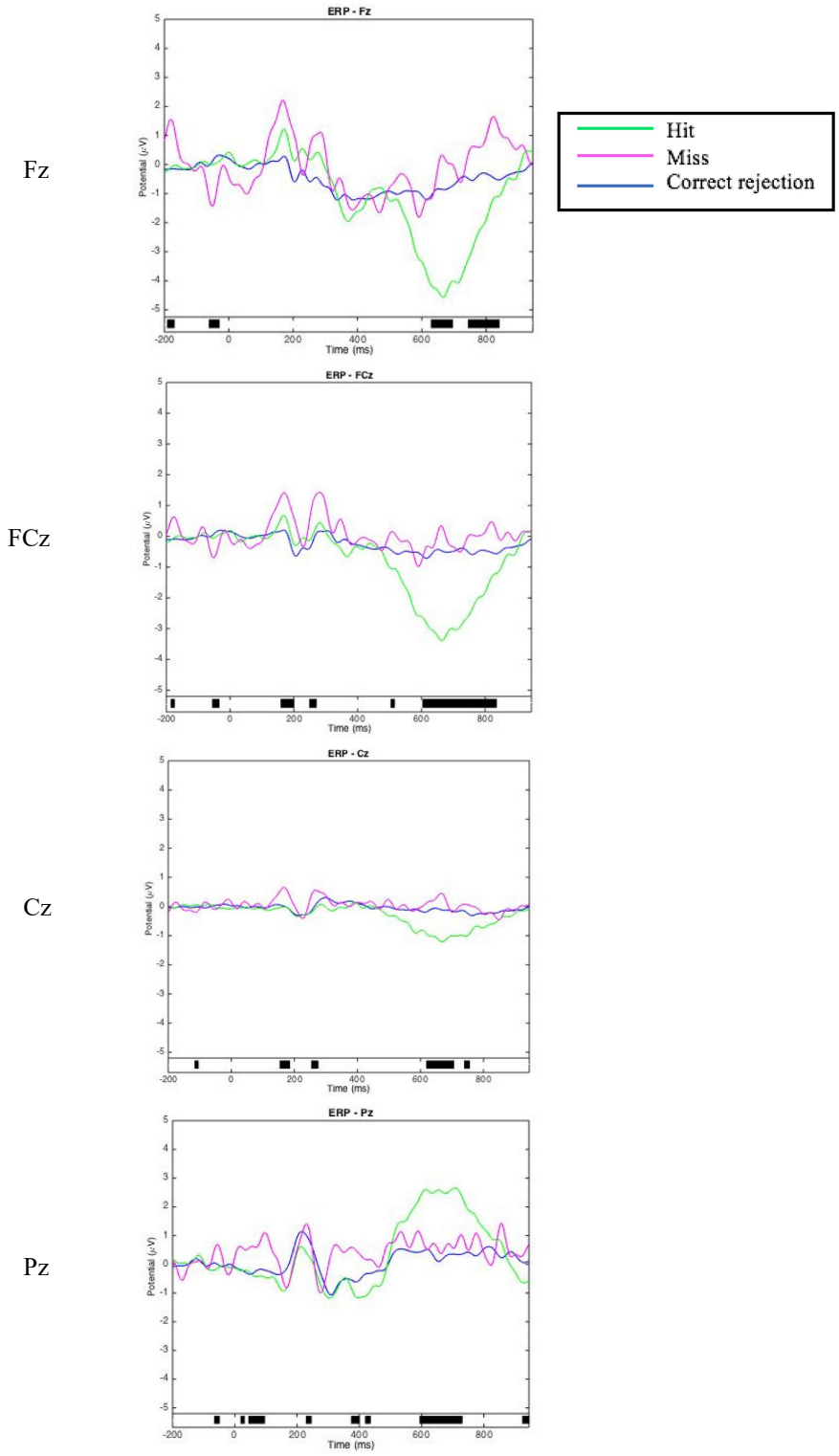


Figure 10.4 Grand average of hits, misses and correct rejections at electrode positions Fz, Cz and Pz, across all participants and all tasks

Table 10.2 lists the averaged classification accuracies of hits vs. correct rejections in different task session. The overall classification accuracy across all task blocks was acceptable and reached 69.37%±8.72. Among all task blocks, the classification accuracies in the Adjective and Calculation task blocks were higher than in other blocks. However, there were no significant differences in classification accuracy across different task blocks.

Table 10.2 Mean classification accuracies of hits vs. correct rejections across all participants in different task blocks (%)

	Music	**News**	**Talk**	**Movie**	**Letter**	**Adjective**	**Calculation**	**Overall**
M	64.89	67.86	66.31	71.11	69.86	73.74	71.80	69.37
SD	8.17	7.89	8.52	4.26	10.95	10.71	6.86	8.72

Due to the very low number of misses, it was not able to classify between hits and misses.

10.2.2 Performance

Table 10.3 lists the average hit rates, miss rates, false alarm rates and correct rejection rates of oddball task under seven different task conditions, across all participants. Averages of these five perception rates were also calculated across all seven task conditions.

Table 10.3 Rates of hits, misses, false alarms, correct rejections, and accuracies under seven different task conditions and average across all these task conditions (%)

	Music	**News**	**Movie**	**Letter**	**Talk**	**Adjective**	**Calculation**	**M**
Workload condition	**Low**	**Low**	**Medium**	**Medium**	**High**	**High**	**High**	
Hit (TP)/Target (Precision)	95.38	84.98	78.64	95.09	80.59	86.43	82.95	86.29
Miss/Target	4.62	15.02	21.36	4.91	19.41	13.57	17.05	13.71
False Alarm/Distractors	2.39	2.69	4.75	1.82	2.08	2.33	1.39	2.49
Correct Rejection (TN)/Distractors (Recall)	97.61	97.31	95.25	98.18	97.92	97.67	98.61	97.51
Accuracy	96.50	91.15	86.95	96.63	89.26	92.05	90.78	91.90

Figure 10.5 shows the average hit rates, miss rates, false alarm rates, correct rejection rates and accuracies under seven different task conditions and their means.

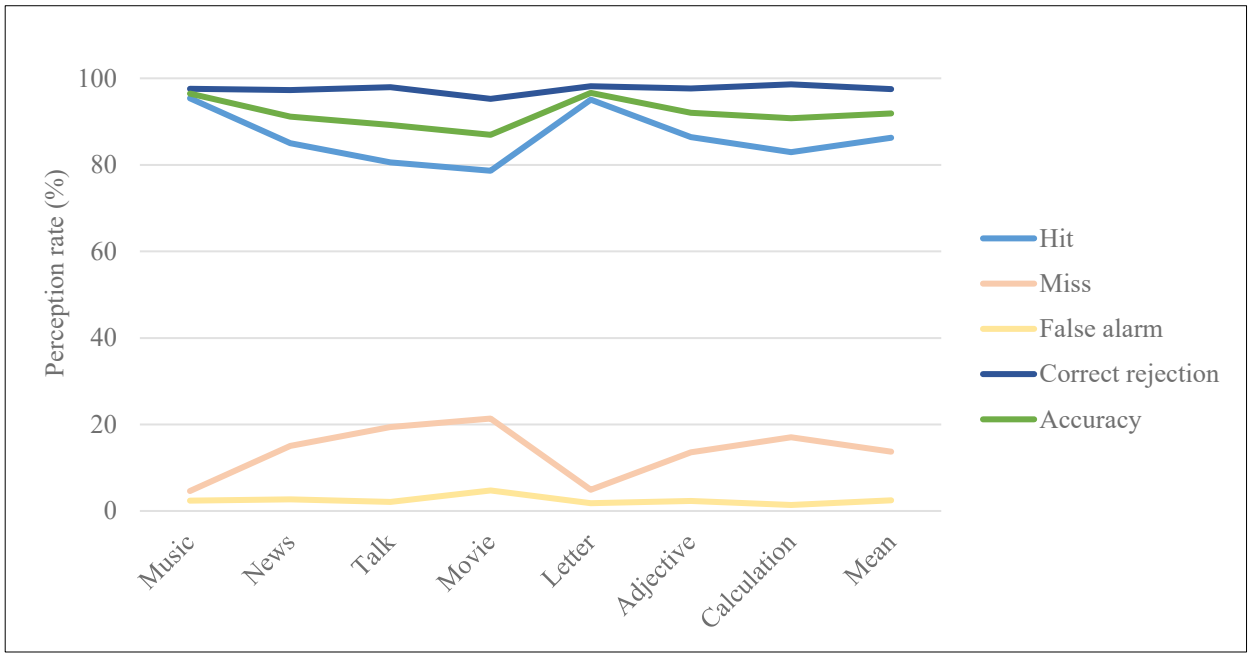


Figure 10.5 Average rates of hits, misses, false alarms, correct rejections and accuracies of all participants under different task conditions

From Table 10.3 and Figure 10.5 we could see, that there were more misses and less hits in most medium and high workload trials, i.e. under talk, movie, adjective and calculation task conditions. On average, there were 13.71% misses of all target stimuli, and only 86.29% of targets were correctly responded. For distractors, almost all of them (97.51%) were correctly ignored.

Figure 10.6 presents miss rates under seven task conditions. There were significantly more misses of oddball task in the talk ($p<0.01$) and movie ($p<0.05$) blocks than in music block. Comparing between oddball task performance in talk and letter task blocks, there was also significant difference ($p<0.01$) in stimuli miss rates.

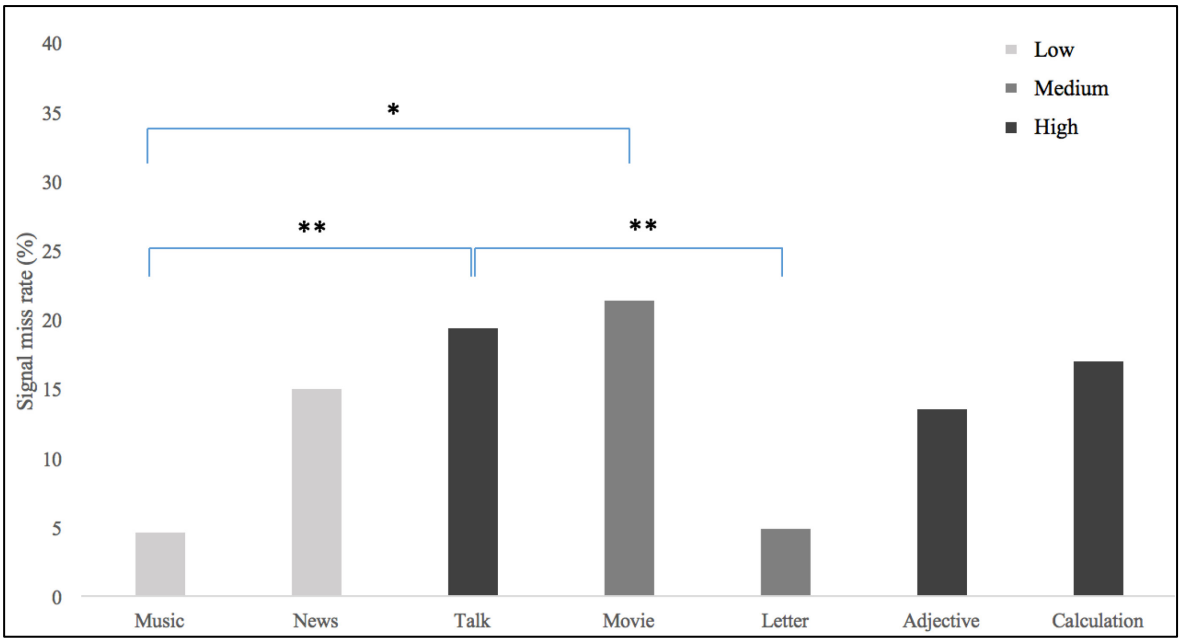


Figure 10.6 Miss rates under seven task conditions. Among them, tasks of different workload levels are also presented in different colors. Paired t-tests were calculated and significances were marked (*p<0.05, **p<0.01)

Table 10.4 lists the reaction times of hits, false alarms and all responses in the oddball task, under seven different task conditions.

Table 10.4 Reaction time of oddball task under different task conditions and their mean

	Music	**News**	**Talks**	**Movie**	**Letter**	**Adjective**	**Calculation**	**M**
RT (Hits)	0.874	0.927	0.948	0.844	0.979	1.171	1.098	0.977
RT (All)	0.874	0.974	0.977	0.948	0.984	1.201	1.180	1.020

Figure 10.7 shows the reaction times of hits and all responses in oddball task under different task conditions. In the music task block, the reaction time of all responses was the shortest. In adjective and calculation task blocks, participants reacted slower than other task blocks. However, paired t-tests results showed that, there was only significant difference in reaction times between music and adjective task blocks (p<0.01).

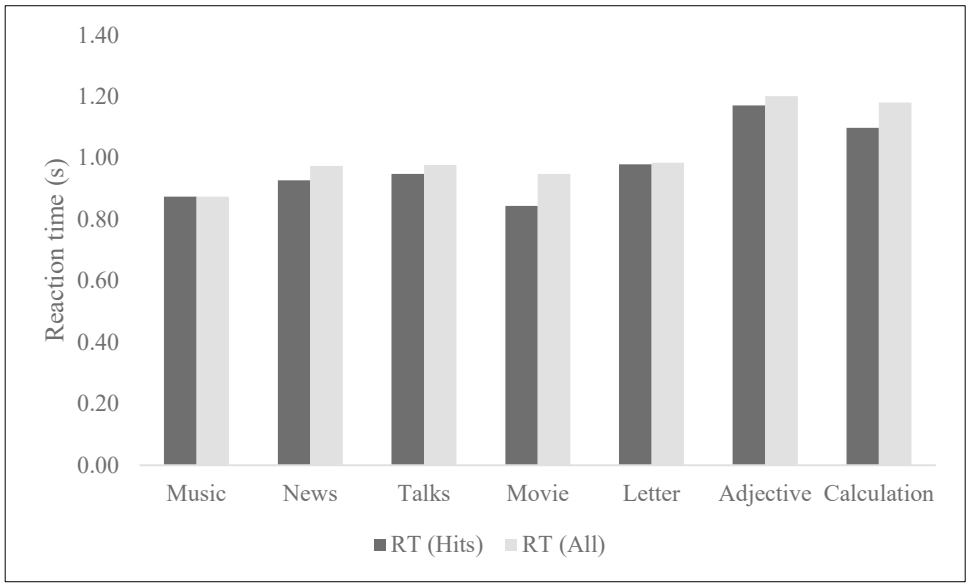


Figure 10.7 Reaction time of oddball task under different task conditions

10.3 Discussion

The main objective of this experiment was to investigate the possibility of detecting signal perception with EEG data, as well as the impact of mental workload on signal perception.

The first aim of this experiment was to implement an EEG-based classifier that recognized the neuronal correlates of auditory stimuli, which should not only be able to classify the presence and absence of a target stimulus based on the EEG signals, but also recognize the correctly responded stimuli (hits). In the oddball paradigm, there were two types of stimuli: targets and distractors. And according to the signal detection theory, there were four classes based on the responses to these two stimulus types, namely hits, misses, false alarms and correct rejections. With the goal of recognizing hits from other stimulus classes, a multi-class classifier appeared to be the optimal solution. However, it turned out that a multi-class classification led to a poor classification rate under all four conditions and had to be rejected, since the number of misses and false alarms was too small and they cannot be grouped as independent classes.

For the reasons mentioned above, the study concentrated on classifiers that could only separate two classes. The first question to be answered was whether a BCI could classify the presence and absence of a signal based on its correlates. As part of this classification analysis, the conditions "hits" and "missed" were combined into a corresponding class and represented the test runs with stimulus presentation (class "target

class"). The conditions resulting from the correct rejections and false alarms were accordingly combined in another class and represented the test runs without any stimulus presentation (class "distractor class").

The offline classification accuracies of the ERPs between targets and distractors show the feasibility to detect signals with ERPs based on EEG data. A similar result could be found in classification between hits and correct rejections. Looking at the ERP curves of corresponding classes, obvious differences in potential between classes can be observed. The illustrations of ERP curves show that a positive wave (PW) occurred in the parietal electrodes and a negative wave (NW) appeared in the frontal electrodes on targets and hits. This further verifies the possibility to differentiate targets and distractors, as well as hits and correct rejections, in regard to ERPs. This indicates that the classifier could recognize the presented stimuli on the basis of the correlates.

Actually, there has been many studies validating the possibility of detecting signal perception with ERPs. Results in this study are consistent with former studies, and furthermore, they build solid basis for future applications in utilizing ERPs to detect signal perception and to indicate vigilance level in more realistic situations.

From performance results under different workload conditions, it can be concluded that workload has an impact on reaction time and detection rate, which could reflect vigilance level. There were several studies on the effect of workload on signal detection rate. For example, Dehais and his colleagues (2016) demonstrated that higher workload leads to a vigilance decrement over time. In our study, data on perception rate and reaction time support a relation between vigilance reduction and workload increase. We could observe this effect when comparing the perception rate in oddball task under different workload conditions. Although not all differences between workload conditions are statistically significant, it is evident from the descriptive results that an increase of workload could lead to an increase in reaction time and a decrease in perception rate (meanwhile also increase in miss rate).

The presence of a linear increase of mean long reaction times corresponds well to classical findings (Williams et al., 1959) and can be explained by the increased number of attentional lapses due to overload. Auditory reaction time in a subsidiary task is a valid estimate for brake reaction time to unexpected obstacles, as shown by Laurell & Lisper (1978). Therefore, it seems reasonable to infer from our data that reaction to the auditory oddball task can to certain extent reflect participants' reactivity to unexpected traffic events and the ability for potential take-over of the vehicle when requested.

The long-term goal of our investigations is the real-time classification of vigilance in a practical scenario. Although reaction times are the most reliable measures for vigilance, we expect to obtain more direct

insights regarding the state of vigilance. This implies the engagement of an online classification of vigilance from EEG data, which is not intrusive and doesn't need any active actions from the participants to intervene the process of other ongoing tasks. Therefore, a further investigation on the online detection of vigilance with EEG would be investigated in the future.

10.4 Summary

Overall, ERPs analysis shows significant differences between targets and distractors, as well as between hits and correct rejections. Classification of these two class pairs indicates acceptable accuracies and it's reasonable to classify between these two class pairs. These results confirmed the viability to detect signal perception from other distractors of the oddball task, which is consistent with results from former studies.

In regard with vigilance under different workload conditions, performance of oddball task varies in accordance with workload levels. Higher mental workload induced by non-driving tasks could result in deteriorated vigilance to detect target stimuli.

Hence, it is of great importance to monitor driver's vigilance level, since they might engage in different tasks while highly automated driving, which could affect their ability to take over control of the vehicle when requested. EEG provides a reliable method to measure the detection of stimuli. In this study, we only investigated the offline classification of ERPs to indicate vigilance. In order to monitor vigilance level in practical scenarios, the next step will be to examine the feasibility of measuring vigilance level with EEG data in real time.

11. Experiment 3 – Online detection of driver's mental workload with EEG in highly automated driving and its impact on take-over control

11.1 Introduction and hypothesis

Although automated driving has been under research for years and there appears to be a consensus that fully automated vehicles will be prevalent on roads in the future, it will also take many more decades till fully automated driving becomes ubiquitous (Shladover et al., 2015). Before full automation is technically achievable, conditional and high automated driving (Level 3 & 4) will be firstly deployed, in which automation is not perfectly reliable and the driver has to take over control of the vehicle in some situations.

In the situation of automated driving, drivers are inclined to be involved in non-driving activities to avoid monotony. This may induce the variation of mental workload and thus impact driver's performance in the take-over situation. As discussed in chapter 3, frequent updates about the drives' states help them better prepare to regain control of the vehicle in a safer manner, and in order to evaluate drivers' capabilities to take over, it is necessary to monitor their mental workload in real time. Whether such updates on drivers' mental workload and warnings when they are overloaded could indeed improve driver's performance in take-over and shorten their take-over time, is a main focus of this study.

Therefore, in this study, we hypothesize that

- it is possible to identify whether the driver is under heavy/low workload in real time across different non-driving activities in a simulated autonomous driving scenario, by means of the former investigated technique based on EEG data.

According to the real-time analysis results of EEG data, the system will give corresponding feedbacks to the drivers if they are under higher workload. Furthermore, we also suppose that

- these feedbacks could help the driver get more ready to take over control of the vehicle and thus shorten the take-over time.

11.2 Methods

11.2.1 Participants

Overall, 8 participants (2 females and 6 males) between the ages of 19-32 (M=23.13, SD=5.28) took part in this experiment. All participants possessed of a valid driving license for 3.75 years on average. 7 of them drove under 10,000 kilometers per year, and the other drove between 10,000 and 20,000 kilometers every year. All of the participants had normal or corrected-to-normal vision, and reported to be free of illness and medication. Caffeine, tobacco, and alcohol were prohibited prior to participating in the experiments. Participants all received a cash payment for their participation.

11.2.2 Experimental design

This experiment featured a three-factorial within-subject design. There were two *independent variables (IVs)*. A concrete illustration of the independent variables can be found in Figure 11.1. The first (IV1) is the objective workload levels of different non-driving tasks, which consists of two levels. In this experiment, participants performed four non-driving tasks, and the workload levels of these tasks were validated by measuring their subjectively perceived workload with questionnaires in the pre-study of Experiment 2 (see Chapter 9.3). The four non-driving tasks were music, talk, adjective, and calculation, and they will be explained in the following part. Among them, Music induces low workload, and the other three induce high workload. The second independent variable (IV2) is whether there was online feedback to the high workload condition. This variable contains two levels: with feedback or no feedback. The performance in high workload tasks without online classification and feedback serves as baseline, and was compared to that of tasks classified as under high workload and with feedback.

The dependent variables in this experiment include the classification results of workload level based on EEG data, one-itemed workload scale for each task, take-over time and subjective ratings on the effect of feedback.

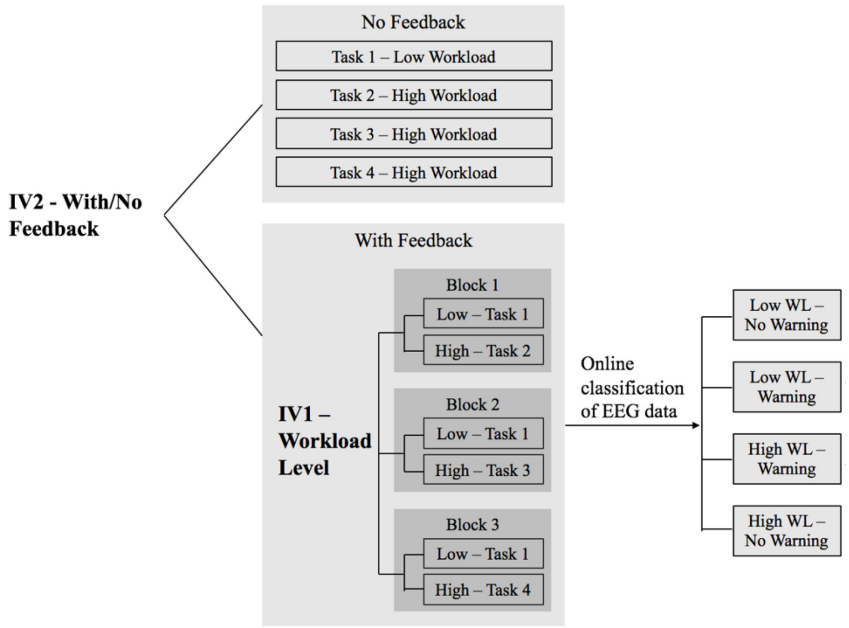


Figure 11.1 Illustration of independent variables in the experimental design

11.2.3 Procedure

Participants were seated in a simulated highly automated driving car, which is driving on a country road at 30 km/h. After the preparation phase of electrode system and the calibration task for workload classifier, which are same as Experiment 1 & 2a, participants were instructed to firstly fulfill a no-feedback session with three different tasks inducing high workload.

Following that, three online feedback blocks were performed, each consisted of the low workload task (music task) and a kind of high workload-inducing task. In each block, the low and high workload tasks each appeared three times, and the sequence of tasks within the block was totally randomized. So was the sequence of three blocks. After 2 minutes of the task start, a judgement of whether the driver was under higher mental workload or not was made, based on the real-time analysis of EEG data. According to the analysis, if the driver was under higher workload, a warning with a reminding sentence was displayed; other wise, no feedback was given to the participant. And the participants could adjust their engagement status based on the appearance of feedback.

After another period of time ranging from 30 to 90 seconds (this time was randomly determined by the system), a take-over request with both a tone (auditory) and a picture on the center stack (visual) was

displayed. Participants were instructed to press the disengagement button on the steering wheel as soon as possible, in order to take over control of the car. Take-over time were recorded. At the end of some random task sessions, participants were instructed to report the answer, in order to ensure their engagement.

Lastly, participants were required to fill a questionnaire on their attitude towards the appearance of feedback, which was used in Yoon and his colleagues' study (2019) based on Van der Laan's questionnaire (1997). The questionnaire consists of four aspects including usefulness, satisfaction, effectiveness and safety, and each question is based on a 7-point Likert scale ranging from -3 (strongly disagree) to +3 (strongly agree).

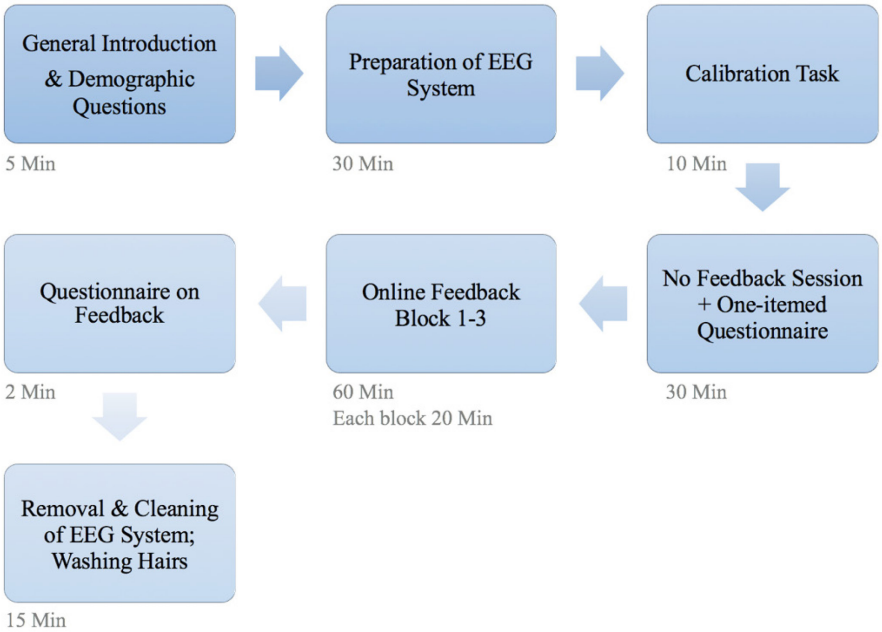


Figure 11.2 General procedure of experiment 3

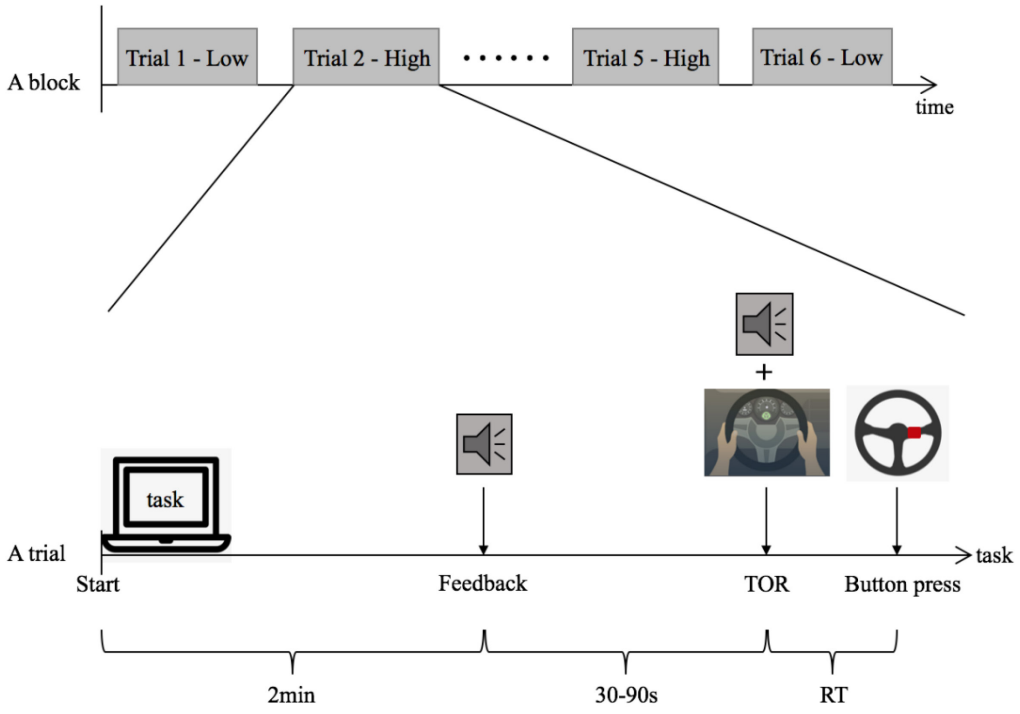


Figure 11.3 Procedure of one online feedback block and one trial in the block

11.2.4 Experiment apparatus

This experiment was conducted in a static driving simulator with steering wheel, gas/brake pedals and other control elements, same as in Experiment 2. The CARLA simulator[11], which is an open-source simulator for autonomous driving research, was used to simulate the autonomous driving environment.

[11] Homepage of simulator: http://carla.org/

Figure 11.4 CARLA simulator environment

EEG data were recorded using 62 active Ag/AgCl electrodes by ESI NeuroScan mounted according to the international 10-20 system. EEG signals were sampled at 1000 Hz and wide-band filtered (0.5-70 Hz).

Different workload-inducing tasks were programmed in Python and displayed with the software PsychoPy on a 17-inch PC screen at the position of Infotainment in vehicle. All auditory materials were played with a loud speaker.

Take-over control was achieved by pressing a button on the steering wheel, as shown in Figure 11.4.

Figure 11.5 Disengagement button on the steering wheel for take-over

Real-time analysis of EEG data was realized with a BCI2000 software and self-programmed function based on MATLAB (version R2014b).

11.2.5 Data analysis

The EEG data analysis methods used were same to those in Experiment 1 & 2 (see Chapter 7 & 9). Firstly, a workload classifier was trained on the calibration task. Then, this classifier was applied in the online feedback blocks. For each task, a real time calculation of classifier outputs was performed for consecutive 1-second epochs, in the time period of 2 minutes. Based on these, we calculated the number of outputs, which were classified into high workload condition, and if this number was greater than 55% of total output number in this task, an auditory warning feedback on the high workload was displayed. The number of truly and wrongly displayed warning feedbacks were also counted and the true positive (high workload task with warning) and true negative (low workload task without warning) rates were calculated.

In addition to the analysis based on EEG data, the take-over time in all blocks were recorded. The average take-over time between no-feedback session and with-feedback session, as well as between high and low workload task session, were compared and their differences were statistically analyzed with paired t-tests.

At last, the ratings of feedback questionnaire on the satisfaction, usefulness, effectiveness and safety aspects were calculated.

The online classification of EEG data was done with the open-source MATLAB-based toolbox BCILAB (version 1.2) the presentation of results in real time was realized with MATLAB (version R2014b). Statistical analysis was performed with SPSS Statistics (version 26.0.0.0).

11.3 Results

11.3.1 Subjective workload measure

Table 11.1 and Figure 11.6 show the mean SEA-scale scores with standard deviations in different tasks over all participants. Results of paired t-tests confirmed that there were significant differences in perceived workload level between music and other tasks (music vs. talk: $t(7)=6.708$, $p<0.001$; music vs. adjective: $t(7)=4.585$, $p=0.003$; music vs. calculation: $t(7)=10.009$, $p<0.001$).

Table 11.1 SEA-scale ratings of four tasks

	Music	**Talk**	**Adjective**	**Calculation**
M	7.41	85.60	88.04	106.16
SD	12.17	35.26	45.47	28.46

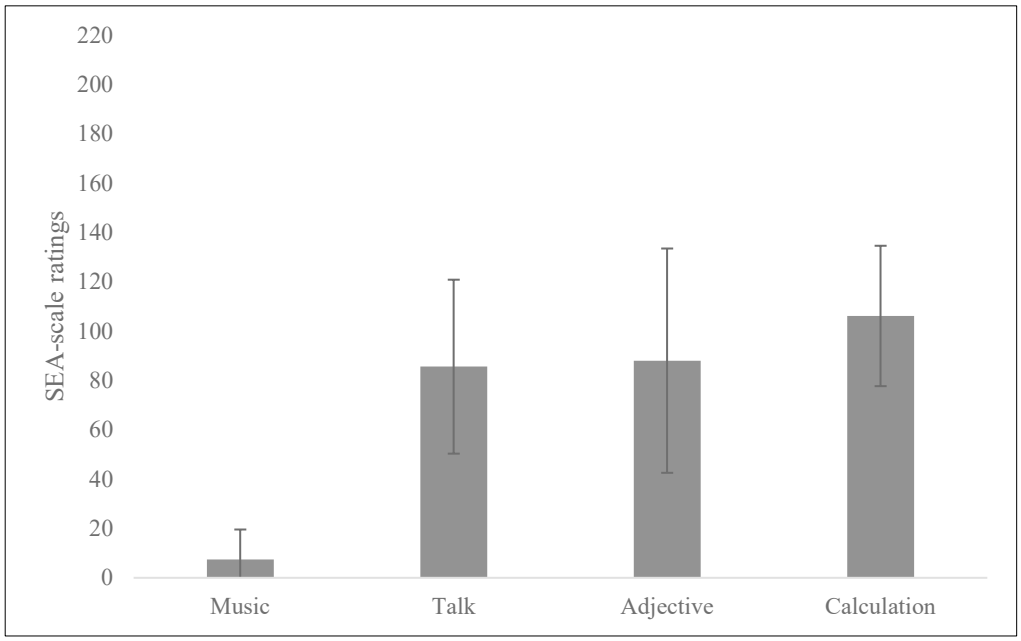


Figure 11.6 SEA-scale ratings of four tasks

11.3.2 Classification results

Table 11.2 lists classification accuracies in the calibration task and those of applying the task-independent classifier in classifying between different high and low workload tasks. The mean estimated classification accuracy over participants of these task pairs reached at least 61.27%.

Table 11.2 Classification accuracy between high and low workload conditions (%)

	Calibration	**Talk vs. Music**	**Text vs. Music**	**Calc vs. Music**
1	65.25	49.12	59.79	61.00
2	69.50	60.96	67.50	62.51
3	74.75	58.48	60.75	57.14
4	65.25	63.99	60.75	66.50
5	87.50	73.09	91.75	84.96
6	63.50	66.34	61.86	65.82
7	59.50	55.45	53.87	63.20
8	72.98	62.72	64.79	62.95
M	69.78	61.27	65.13	65.51
SD	8.74	7.21	11.46	8.37

Table 11.3 lists the feedback appearance count and rate under different workload conditions. In one trial lasting around three minutes, if the system detected participant's higher workload, a warning was accordingly given, and vice verse. In total, in about 90% of trials, proper corresponding feedback were correctly given.

Table 11.3 Feedback appearance rate

	Low Workload		**High Workload**			
Participant	**No Warning (TN)**	**Warning (FA)**	**No Warning (Miss)**	**Warning (TP)**	**TN-rate (%)**	**TP-rate (%)**
1	10	2	1	11	83.33	91.67
2	11	1	2	11	91.67	84.62
3	9	3	2	10	75.00	83.33
4	12	0	1	12	100.00	92.31
5	11	1	2	11	91.67	84.62
6	12	0	2	12	100.00	85.71
7	12	0	0	12	100.00	100.00
8	11	1	1	12	91.67	92.31
M					91.67	89.32
SD					8.91	5.74

11.3.3 Take-over time

Table 11.4 and Figure 11.7 present averaged take-over time of high and low workload tasks across all participants and repetitions of each tasks. Participants used longer time to react to the take-over request when undertaking higher workload tasks. T-tests were performed on each task pairs of high and low workload tasks and their differences were calculated. Results from t-tests showed significant differences between all high and low workload tasks in take-over time.

Table 11.4 Take-over time (s) of high and low workload tasks and their comparison results

	Music vs. Talk	**Music vs. Adjective**	**Music vs. Calculation**
High WL	1.53	1.61	1.66
Low WL	1.38	1.32	1.32
t-test results	t(23)=3.046, p=0.019	t(23)=2.570, p=0.037	t(23)=6.596, p=0.000

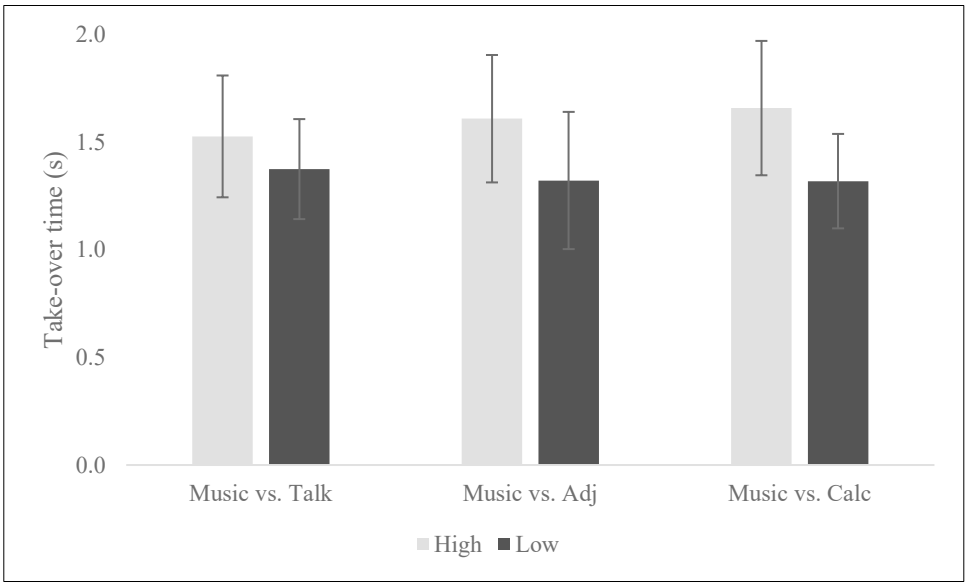


Figure 11.7 Take-over time of high and low workload tasks

Table 11.5 and Figure 11.8 show averaged take-over time of high workload tasks in no feedback session and online feedback session, across all participants and repetitions of each tasks. The averaged take-over times were longer in no feedback session, in comparison to with online feedback session. T-tests were performed and differences were calculated to indicate significant differences in take-over time between no feedback session and online feedback session for all three kinds of tasks.

Table 11.5 Take-over time (s) of high workload tasks in with & no feedback sessions and their comparison results

	Talk	**Adjective**	**Calculation**
No feedback session	1.76	1.85	1.83
Feedback	1.52	1.61	1.65
t-test results	t(23)=6.271, p=0.000	t(23)=3.113, p=0.005	t(23)=2.441, p=0.023

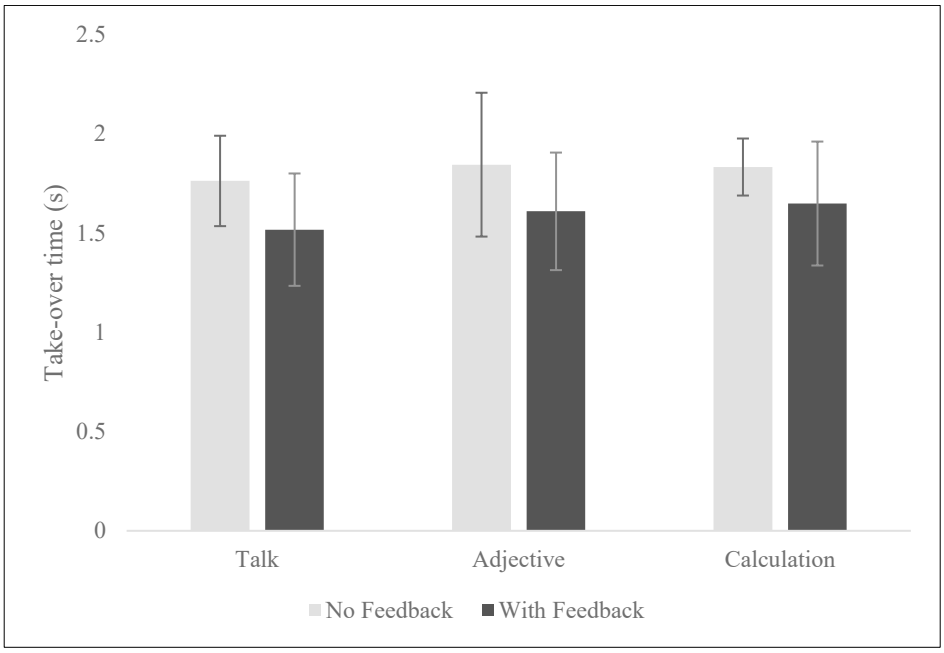


Figure 11.8 Take-over time of high workload tasks in with & no feedback sessions

11.3.4 Feedback ratings

Table 11.6 and Figure 11.9 present averaged subjective ratings on the appearance of feedback warnings. Scores showed positive ratings (all above 1) in all four aspects.

Table 11.6 Subjective ratings on the appearance of feedback warnings (7-point scores ranging from -3 to +3)

	Satisfaction	**Usefulness**	**Effectiveness**	**Safety**
M	1.50	1.75	1.88	1.75
SD	1.69	2.05	1.64	1.39

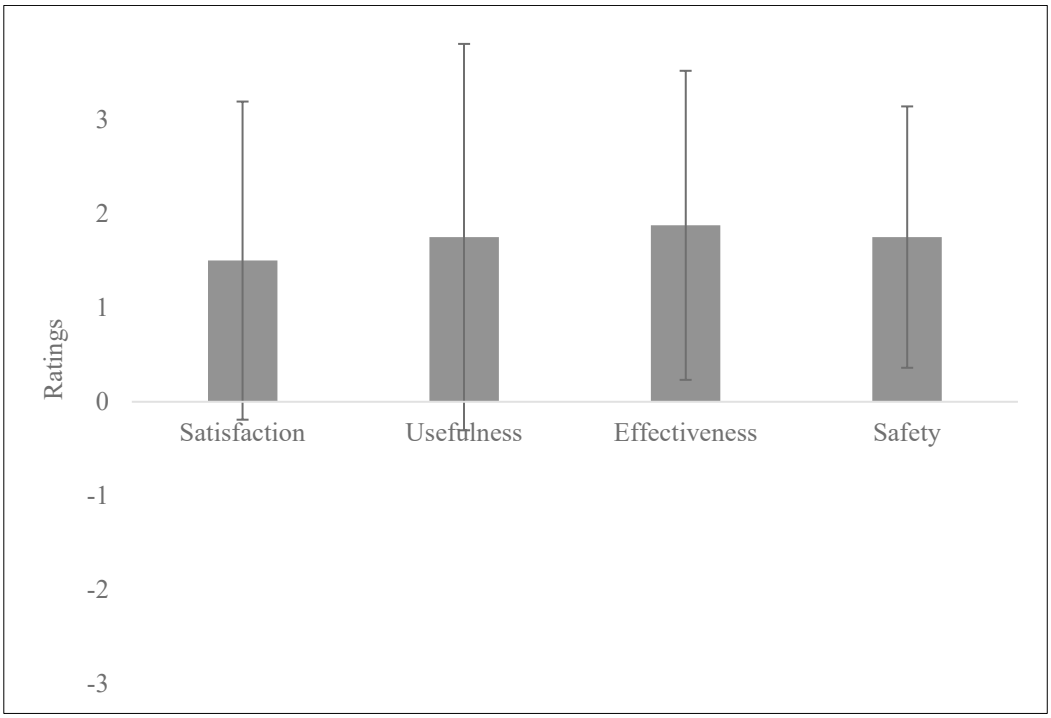


Figure 11.9 Subjective ratings on the appearance of feedback warnings[12]

[12] The subjective rating data are not normal distributed here.

11.4 Discussion

The aim of this study was to investigate the applicability of task-independent classifier in practical workload detection scenario in real time and the effect of feedback appearance on take-over performance. We also investigated the subjective perception of feedback in satisfaction, usefulness, effectiveness and safety aspects.

Subjective ratings on perceived mental workload confirmed the results in previous study (Experiment 2) and work as a basis for the setup of workload conditions (high/low) in this experiment.

Classification accuracies between high and low workload conditions when applying the task-independent workload classifier are analog to Experiment 2, and thus verifies the efficiency of this task-independent workload classifier. Combined with the feedback appearance rates, it was further proven that in most cases, it was able to classify the workload levels of different practical tasks into correct classes. Albeit a few wrong appearances of feedback warning in the low workload conditions, there were appropriate feedbacks in most blocks according to the estimated workload level.

Take-over times also showed significant differences between high and low workload conditions. This means that higher workload non-driving tasks could cause a lower situation awareness and thus a decrement in react ability to take-over compared with low workload tasks. Drivers occupying themselves with much easier non-driving tasks show better take-over behavior due to the possibility of observing the surrounding traffic situation and perceiving the take-over request in time. It is thus demonstrated that high mental workload does have a negative effect on driver's performance in the take-over situation.

Results on take-over time showed that these feedbacks were effective in shortening drivers' take-over time under high mental workload, which is a critical factor in the situation of take-over control. The take-over time in the feedback sessions was significantly shorter than that in no feedback sessions. This implies that the participants were able to react faster to the take-over request if they were before warned of their relatively higher mental workload and adjust their involvement in the task. As demonstrated by many previous studies (e.g. Gold et al., 2015) and according to results from Experiment 2, higher mental workload may lead to deteriorated ability of driver to react to the displayed take-over stimulus and result in decreased vigilance level. Under the condition of with feedback, participants were informed of their too high mental workload level and could adjust their status to reach a moderate level of workload. This could help them be more ready for the take-over request.

When further observing the reaction times under different conditions, the averages of reaction time are all under 2 seconds, with maximum under 3 seconds. All the reaction times of all participants are far below 7 seconds, which is the suggested take-over time by many researches (e.g. Gold et al., 2013; Radlmayr et al., 2014). However, the take-over time measured in this experiment was through a button press on the steering wheel to enable the switch of driving mode. Here we didn't inspect driver's behavior of braking or accelerating, and steering the wheel, which are components of the actual maneuver in driving. According to Gold and his colleagues' study (2013), the time of a real intervention is a bit higher than the hands on steering wheel time. Thus, the influence of different conditions on these behaviors should also be investigated in the future.

In regard with subjective ratings on the appearance of feedback warnings, participants' overall attitudes were positive. That means, they were pro these feedback warnings. However, some participants were not very satisfied with them and found them not so useful, although the ratings on effectiveness and safety were relatively higher. This might be due to the appearance form of the feedback warning. In this experiment, the feedback was just an auditory displayed sentence. Whether it should be performed in a visual way as well and the specific content of the sentence, are still not investigated and should be studied in the next step.

11.5 Summary

This study investigated the application of EEG in detecting mental workload level in real time, and the influence of feedback to overload on the take-over process during highly automated driving. Results indicate the feasibility of detection mental workload with EEG, by means of applying a simplified task-independent workload classifier in different non-driving tasks. Classification accuracies and feedback appearance rates both confirmed the viability of this classifier. Take-over time under different mental workload conditions demonstrated the impact of workload on driver's performance in take-over. Results on take-over time also show a strong influence of the appearance of feedback warnings when overloaded on take-over time and quality.

Furthermore, this experiment is limited due to its conduction in a driving simulator and a low number of participants. An enlarged sample size and the conduction in a real-life scenario are needed to examine the effect more precisely and more specifically.

12.Conclusions and outlook

12.1 Conclusions

In contrast to manual driving, highly automated driving may relieve the driver from active driving and observing activities. However, undertaking different non-driving tasks while highly automated driving could lead to variations in driver's mental workload and vigilance levels. This has great influence on driver's ability and performance to take over control of the vehicle when requested. Therefore, the overall aim of the thesis was to investigate methods to measure drivers' mental workload and vigilance non-intrusively as that potentially improves driver's ability for take-over request in urgent situations. To achieve this, three separate studies were carried out.

In the first study, a simplified task-independent workload classifier based on EEG data was investigated in a well-controlled laboratory. Individual classifier was trained on a standardized workload-inducing task in the calibration phase, and then applied to data from five other cognitive tasks each inducing two workload levels to classify different workload levels. Results indicated that a direct transfer across tasks was not optimal to separate high and low workload levels with acceptable accuracy, however, there were indeed significant differences in the classifier outputs between conditions of five tasks after applying this task-independent classifier. It was implied to look at the classifier output on a continuous scale as an indication of the actual workload level, rather than just take the binary classification accuracy into consideration. Moreover, it was possible to calibrate a single classifier on a standardized workload-inducing task and apply this classifier to other tasks to obtain meaningful results on workload levels. This enables a simplified implementation of real-time and continuous workload detection with EEG for future applications.

The second study was based on the first study and extended the application of the task-independent workload classifier into a more realistic simulated driving situation. At the same time, this study also investigated the detection of vigilance level under workload in such simulated highly automated driving situation. Results showed that the simplified task-independent workload classifier was able to indicate mental workload with classifier outputs, which can significantly differentiate between classes. This task-independent workload classifier was demonstrated as an applicable tool in more realistic and practical scenarios. Moreover, ERP analysis results of oddball task indicated the feasibility of detecting signal perception with EEG and thus inferring driver's vigilance level based on this. Besides, variations in target stimuli detection under different workload conditions confirmed the influence of workload on driver's vigilance level to detect signals.

The third study was an online application of the in study one and two investigated method for workload detection. This study also examined the impact of feedback appearance on driver's take-over performance

in a simulated highly automated driving scenario. Results indicated that it was possible to identify driver's workload level in real time with the task-independent workload classifier. Furthermore, online feedback of heavy workload was proven to be useful in improving driver's take-over performance, as the take-over times in with-feedback sessions were significantly shorter than those in baseline. In addition, subjective ratings on feedback appearance also showed its positive effect.

Based on these three studies, we could conclude that, it is possible to measure driver's mental workload and vigilance levels during highly automated driving by decoding brain signals. Moreover, we investigated a simplified mental workload detection method to classify mental workload, which is independent on the task in the calibration session. Our studies proved the hypothesis, that it is able to calibrate a workload classifier on a standard workload-inducing task and apply on other different testing tasks. This task-independent workload classifier greatly simplifies the training session and shortens calibration time. Furthermore, we also successfully applied this method in an online study and verified its effectiveness in mental workload detection and warning during highly automated driving. Apart from mental workload measurement, the use of brain signals to detect driver's readiness to stimulus and vigilance level was validated. Besides, results also confirmed that mental workload could influence driver's vigilance level.

12.2 Outlook

In this work, the studies were carried out in a highly automated driving scenario and an application of driver state monitoring was investigated, which might be a reality in a few years. Based on the literature research, the studies in this thesis are the first of this kind to examine driver's mental states continuously and non-intrusively in the context of highly automated driving. In future, the outcomes of these studies could contribute to the design of driver-vehicle-interaction in highly autonomous driving by identifying driver's mental states during mode transition, and thus ensuring the safety and user experience of drivers.

However, the three studies in this work were conducted either in a well-controlled laboratory or in a driving simulator, and such settings are far-off from the real traffic situations. Hence, a study in real traffic situation should be carried out in the future. Nevertheless, the implementation of Level 3 autonomous vehicle is still under investigation and test, and currently there's still no real Level 3 autonomous vehicle driving on the road. Application of driver state monitoring investigated in this thesis in real traffic situation needs longer time to realize. Probably the first step in the future would be an investigation in a real car only driving in the limited test field.

Besides, a larger number of participants would be ideal here in order to test enough drivers with online feedback. Although the results showed significant differences, the sample size of participants in the online study of Experiment 3 was small, and we still could not make a solid conclusion that the results were robust and stable enough to be transferred to other experimental settings or replicated in other situations.

Moreover, since we aim to apply the EEG system to monitor driver state in a realistic environment and probably in a real autonomous driving vehicle in the future, such an EEG system with a large amount of electrodes and requiring a long preparation to apply the conductive gel is not an optimal choice. To avoid the cumbersome preparation of traditional EEG systems, an alternative EEG system with dry electrodes was investigated (e.g. Zander et al., 2011; Chi et al., 2011). The electrodes of this system were adapted to establish a direct contact with the skin without the use of conductive gel. In contrast to the traditional EEG systems, this dry electrodes system requires less preparation time and increases comfort for the wearer. Based on the evaluation of BCI classification accuracy, amplitude and temporal structures of ERPs as well as features in the frequency domain, the applicability of dry electrodes system in the context of a running vehicle was validated (Zander et al., 2017), thus demonstrating the feasibility of applying a dry electrode system in autonomous driving. Therefore, in the future, monitoring driver's mental states with a much more portable dry electrode system should be further investigated.

In addition, as the overall goal is to develop a comprehensive system to monitor different aspects of driver's mental states, it is necessary to combine with other physiological methods to get a full scope, such as applying eye-tracking system to detect driver's drowsiness. The merge with other information could improve the detection of driver states' variations, so that the driver could get more ready for a safe transition between driving modes.

References

Abe, T., Nonomura, T., Komada, Y., Asaoka, S., Sasai, T., Ueno, A., & Inoue, Y. (2011). Detecting deteriorated vigilance using percentage of eyelid closure time during behavioral maintenance of wakefulness tests. *International Journal of Psychophysiology*, *82*(3), 269-274.

Anderson, J. M., Kalra, N., Stanley, K. D., Sorensen, P., Samaras, C., & Oluwatola, O. (2016). Autonomous vehicle technology: a guide for policymakers. Santa Monica, CA: RAND Corporation.

Ang, K. K., Chin, Z. Y., Zhang, H., & Guan, C. (2008, June). Filter bank common spatial pattern (FBCSP) in brain-computer interface. In *2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence)* (pp. 2390-2397). IEEE.

Bainbridge, L. (1983). Ironies of automation. In *Analysis, design and evaluation of man–machine systems* (pp. 129-135). Pergamon.

Bansal, D., & Mahajan, R. (2019). *EEG-Based Brain-Computer Interfaces: Cognitive Analysis and Control Applications*. Academic Press.

Baranski, J. V. (2007). Fatigue, sleep loss, and confidence in judgment. *Journal of Experimental Psychology: Applied*, *13*(4), 182.

Beatty, J. (1982). Task-evoked pupillary responses, processing load, and the structure of processing resources. *Psychological bulletin*, *91*(2), 276.

Belz, S. M., Robinson, G. S., & Casali, J. G. (2004). Temporal separation and self-rating of alertness as indicators of driver fatigue in commercial motor vehicle operators. *Human factors*, *46*(1), 154-169.

Berger, H. (1929). Über das Elektroenzephalogramm des Menschen. *European Archives of Psychiatry and Clinical Neuroscience*, 87(1), 527–570.

Berka, C., Levendowski, D. J., Lumicao, M. N., Yau, A., Davis, G., Zivkovic, V. T., ... & Craven, P. L. (2007). EEG correlates of task engagement and mental workload in vigilance, learning, and memory tasks. *Aviation, space, and environmental medicine*, *78*(5), B231-B244.

Billings, C. E. (1991). *Human-centered aircraft automation: A concept and guidelines* (Vol. 103885). National Aeronautics and Space Administration, Ames Research Center.

Bimbraw, K. (2015, July). Autonomous cars: Past, present and future a review of the developments in the last century, the present scenario and the expected future of autonomous vehicle technology. In *2015 12th international conference on informatics in control, automation and robotics (ICINCO)* (Vol. 1, pp. 191-198). IEEE.

Birbaumer, N., & Schmidt, R. F. (1996). Biologische Psychologie. 3. komplett überarbeitete Auflage Berlin. *Heidelberg/New York*.

Blankertz, B., Lemm, S., Treder, M., Haufe, S., & Müller, K. R. (2011). Single-trial analysis and classification of ERP components—a tutorial. *NeuroImage*, *56*(2), 814-825.

Blankertz, B., Tomioka, R., Lemm, S., Kawanabe, M., & Muller, K. R. (2007). Optimizing spatial filters for robust EEG single-trial analysis. *IEEE Signal processing magazine*, *25*(1), 41-56.

Bonnefond, A., Doignon-Camus, N., Touzalin-Chretien, P., & Dufour, A. (2010). Vigilance and intrinsic maintenance of alert state: An ERP study. *Behavioural brain research*, *211*(2), 185-190.

Boostani, R., & Moradi, M. H. (2004). A new approach in the BCI research based on fractal dimension as feature and Adaboost as classifier. *Journal of Neural Engineering*, *1*(4), 212.

Bornemann, E. (1942). Untersuchungen über den Grad der geistigen Beanspruchung. *Arbeitsphysiologie*, *12*(2), 142-172.

Boye, A. T., Kristiansen, U. Q., Billinger, M., do Nascimento, O. F., & Farina, D. (2008). Identification of movement-related cortical potentials with optimized spatial filtering and principal component analysis. *Biomedical Signal Processing and Control*, *3*(4), 300-304.

Bronzino, J. D. (1995). Principles of electroencephalography. *The biomedical engineering handbook*, *1*.

Brookhuis, K. A., De Waard, D., & Janssen, W. H. (2001). Behavioural impacts of advanced driver assistance systems–an overview. *European Journal of Transport and Infrastructure Research*, *1*(3), 245-253.

Brookings, J. B., Wilson, G. F., & Swain, C. R. (1996). Psychophysiological responses to changes in workload during simulated air traffic control. *Biological psychology*, *42*(3), 361-377.

Brouwer, A. M., Hogervorst, M. A., Van Erp, J. B., Heffelaar, T., Zimmerman, P. H., & Oostenveld, R. (2012). Estimating workload using EEG spectral power and ERPs in the n-back task. *Journal of neural*

engineering, *9*(4), 045008.

Brown, I. D., & Poulton, E. C. (1961). Measuring the spare 'mental capacity'of car drivers by a subsidiary task. *Ergonomics*, *4*(1), 35-40.

Brown, T., Lee, J., Schwarz, C., Fiorentino, D., & McDonald, A. (2014). *Assessing the feasibility of vehicle-based sensors to detect drowsy driving* (No. DOT HS 811 886).

Brysbaert, M., Buchmeier, M., Conrad, M., Jacobs, A. M., Bölte, J., & Böhl, A. (2011). The word frequency effect. *Experimental psychology*.

Buld, S., Krüger, H. P., Hoffmann, S., Kaussner, A., Tietze, H., & Totzke, I. (2002). Wirkungen von Assistenz und Automation auf Fahrerzustand und Fahrsicherheit. *Abschlussbericht BMBF*, *19*, 9812.

Byers, J. C. (1989). Traditional and raw task load index (TLX) correlations: Are paired comparisons necessary?. *Advances in Industrial Erfonomics and Safety l: Taylor and Francis.*

Byrne, E. A., & Parasuraman, R. (1996). Psychophysiology and adaptive automation. Biological Psychology, 42, 249-268.

Cain, B. (2007). A review of the mental workload literature: DTIC Document.

Chaouachi, M., Jraidi, I., & Frasson, C. (2011, July). Modeling mental workload using EEG features for intelligent systems. In *International Conference on User Modeling, Adaptation, and Personalization* (pp. 50-61). Springer, Berlin, Heidelberg.

Chi, Y. M., Wang, Y. T., Wang, Y., Maier, C., Jung, T. P., & Cauwenberghs, G. (2011). Dry and noncontact EEG sensors for mobile brain–computer interfaces. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, *20*(2), 228-235.

Chiappa, S., & Bengio, S. (2003). *HMM and IOHMM modeling of EEG rhythms for asynchronous BCI systems* (No. REP_WORK). IDIAP.

Chira-Chavala, T., & Yoo, S. M. (1994). Potential safety benefits of intelligent cruise control systems. *Accident Analysis & Prevention*, *26*(2), 135-146.

Cicchino, J. B. (2017). Effectiveness of forward collision warning and autonomous emergency braking systems in reducing front-to-rear crash rates. *Accident Analysis & Prevention*, *99*, 142-152.

Czisch, M., Wehrle, R., Harsay, H. A., Wetter, T. C., Holsboer, F., Sämann, P. G., & Drummond, S. (2012). On the need of objective vigilance monitoring: effects of sleep loss on target detection and task-negative activity using combined EEG/fMRI. *Frontiers in neurology*, *3*, 67.

Damböck, D., Farid, M., Tönert, L., & Bengler, K. (2012). Übernahmezeiten beim hochautomatisierten Autofahren. 5. Tagung Fahrerassistenz 2012. *München. Germany*.

Dehais, F., Roy, R. N., Gateau, T., & Scannella, S. (2016, July). Auditory alarm misperception in the cockpit: an EEG study of inattentional deafness. In *International Conference on Augmented Cognition* (pp. 177-187). Springer, Cham.

De Waard, D., & te Groningen, R. (1996). *The measurement of drivers' mental workload*. Netherlands: Groningen University, Traffic Research Center.

De Winter, J. C., Happee, R., Martens, M. H., & Stanton, N. A. (2014). Effects of adaptive cruise control and highly automated driving on workload and situation awareness: A review of the empirical evidence. *Transportation research part F: traffic psychology and behaviour*, *27*, 196-217.

Dickmanns, E. D. (2002, June). The development of machine vision for road vehicles in the last decade. In *Intelligent Vehicle Symposium, 2002. IEEE* (Vol. 1, pp. 268-281). IEEE.

Dinges, D. F. (1995). An overview of sleepiness and accidents. *Journal of sleep research*, *4*, 4-14.

Dooley, C. (2009). The impact of meditative practices on physiology and neurology: a review of the literature.

Dornhege, G., Blankertz, B., Krauledat, M., Losch, F., Curio, G., & Muller, K. R. (2006). Combined optimization of spatial and temporal filters for improving brain-computer interfacing. *IEEE transactions on biomedical engineering*, *53*(11), 2274-2281.

Duda, R. O., Hart, P. E., & Stork, D. G. (2012). *Pattern classification*. John Wiley & Sons. Nunez et al., 1994

Driving, A. (2014). Levels of driving automation are defined in new SAE international standard J3016: 2014. *SAE International: Warrendale, PA, USA*.

Eggemeier, F. T., Wilson, G. F., Kramer, A. F., & Damos, D. L. (1991). Workload assessment in multi-task environments. *Multiple-task performance*, 207-216.

Eilers, K., Nachreiner, F., & Hänecke, K. (1986). Entwicklung und Überprüfung einer Skala zur Erfassung subjektiv erlebter Anstrengung. *Zeitschrift für Arbeitswissenschaft*, (4), 214-224.

Endsley, M. R., & Kiris, E. O. (1995). The out-of-the-loop performance problem and level of control in automation. *Human factors*, *37*(2), 381-394.

Endsley, M. R. (1995). Measurement of situation awareness in dynamic systems. *Human factors*, *37*(1), 65-84.

Eulitz, C., Maess, B., Pantev, C., Friederici, A. D., Feige, B., & Elbert, T. (1996). Oscillatory neuromagnetic activity induced by language and non-language stimuli. *Cognitive brain research*, *4*(2), 121-132.

Fatourechi, M., Bashashati, A., Ward, R. K., & Birch, G. E. (2007). EMG and EOG artifacts in brain computer interface systems: A survey. *Clinical neurophysiology*, *118*(3), 480-494.

Fildes, B., Keall, M., Bos, N., Lie, A., Page, Y., Pastor, C., ... & Tingvall, C. (2015). Effectiveness of low speed autonomous emergency braking in real-world rear-end crashes. *Accident Analysis & Prevention*, *81*, 24-29.

Flemisch, F. O., Adams, C. A., Conway, S. R., Goodrich, K. H., Palmer, M. T., & Schutte, P. C. (2003). The H-Metaphor as a guideline for vehicle automation and interaction.

Flexer, A., Bauer, H., Pripfl, J., & Dorffner, G. (2005). Using ICA for removal of ocular artifacts in EEG recorded from blind subjects. *Neural Networks*, *18*(7), 998-1005.

Fournier, L. R., Wilson, G. F., & Swain, C. R. (1999). Electrophysiological, behavioral, and subjective indexes of workload when performing multiple tasks: manipulations of task difficulty and training. *International Journal of Psychophysiology*, *31*(2), 129-145.

Freeman, F. G., Mikulka, P. J., Scerbo, M. W., & Scott, L. (2004). An evaluation of an adaptive automation system using a cognitive vigilance task. *Biological psychology*, *67*(3), 283-297.

Friedman, J. H. (1989). Regularized discriminant analysis. *Journal of the American statistical association*, *84*(405), 165-175.

Galéra, C., Orriols, L., M'Bailara, K., Laborey, M., Contrand, B., Ribéreau-Gayon, R., ... & Maury, B. (2012). Mind wandering and driving: responsibility case-control study. *Bmj*, *345*, e8105.

Gasser, T. M., Arzt, C., Ayoubi, M., Bartels, A., Bürkle, L., Eier, J., ... & Lotz, C. (2012). Rechtsfolgen zunehmender fahrzeugautomatisierung. *Berichte der Bundesanstalt für Straßenwesen. Unterreihe Fahrzeugtechnik*, (83).

Gevins, A., & Smith, M. E. (2000). Neurophysiological measures of working memory and individual differences in cognitive ability and cognitive style. *Cerebral cortex*, *10*(9), 829-839.

Gevins, A. S., & Smith, M. E. (2003). Neurophysiological measures of cognitive workload during human-computer interaction. *Theoretical Issues in Ergonomics Science, 4*, 113–131.

Gevins, A., Smith, M. E., Leong, H., McEvoy, L., Whitfield, S., Du, R., & Rush, G. (1998). Monitoring working memory load during computer-based tasks with EEG pattern recognition methods. *Human factors*, *40*(1), 79-91.

Gevins, A., Smith, M. E., McEvoy, L., & Yu, D. (1997). High-resolution EEG mapping of cortical activation related to working memory: effects of task difficulty, type of processing, and practice. *Cerebral cortex (New York, NY: 1991)*, *7*(4), 374-385.

Gevins, A. S., Bressler, S. L., Cutillo, B. A., Illes, J., Miller, J. C., Stern, J., & Jex, H. R. (1990). Effects of prolonged mental work on functional brain topography. *Electroencephalography and Clinical Neurophysiology*, *76*(4), 339-350.

Gold, C., Körber, M., Hohenberger, C., Lechner, D., & Bengler, K. (2015). Trust in automation–Before and after the experience of take-over scenarios in a highly automated vehicle. *Procedia Manufacturing*, *3*, 3025-3032.

Gonçalves, J., Olaverri-Monreal, C., & Bengler, K. (2015, September). Driver Capability Monitoring in Highly Automated Driving: from state to capability monitoring. In *2015 IEEE 18th International Conference on Intelligent Transportation Systems* (pp. 2329-2334). IEEE.

Graw, P., Kräuchi, K., Knoblauch, V., Wirz-Justice, A., & Cajochen, C. (2004). Circadian and wake-dependent modulation of fastest and slowest reaction times during the psychomotor vigilance task. *Physiology & behavior*, *80*(5), 695-701.

Gundel, A., & Wilson, G. F. (1992). Topographical changes in the ongoing EEG related to the difficulty of mental tasks. *Brain topography*, *5*(1), 17-25.

Hagemann, K. (2008). The alpha band as an electrophysiological indicator for internalized attention and high mental workload in real traffic driving. *PhD, Mathematics and Natural Sciences, Heinrich-Heine University of Dusseldorf, Dusseldorf*.

Hämäläinen, M., Hari, R., Ilmoniemi, R. J., Knuutila, J., & Lounasmaa, O. V. (1993). Magnetoencephalography—theory, instrumentation, and applications to noninvasive studies of the working human brain. *Reviews of modern Physics*, *65*(2), 413.

Harper, C. D., Hendrickson, C. T., Mangones, S., & Samaras, C. (2016). Estimating potential increases in travel with autonomous vehicles for the non-driving, elderly and people with travel-restrictive medical conditions. *Transportation research part C: emerging technologies*, *72*, 1-9.

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In *Advances in psychology* (Vol. 52, pp. 139-183). North-Holland.

Helton, W. S., & Russell, P. N. (2012). Brief mental breaks and content-free cues may not keep you focused. *Experimental brain research*, *219*(1), 37-46.

Hess, E. H., & Polt, J. M. (1964). Pupil size in relation to mental activity during simple problem-solving. *Science*, *143*(3611), 1190-1192.

Hogervorst, M. A., Brouwer, A. M., & Van Erp, J. B. (2014). Combining and comparing EEG, peripheral physiology and eye-related measures for the assessment of mental workload. *Frontiers in neuroscience*, *8*, 322.

Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H., & Van de Weijer, J. (2011). *Eye tracking: A comprehensive guide to methods and measures*. OUP Oxford.

Horne, J. A., & Baulk, S. D. (2004). Awareness of sleepiness when driving. *Psychophysiology*, *41*(1), 161-165.

Hu, M., Li, J., Li, G., Tang, X., & Ding, Q. (2006, July). Classification of normal and hypoxia EEG based on approximate entropy and welch power-spectral-density. In *The 2006 IEEE International Joint Conference on Neural Network Proceedings* (pp. 3218-3222). IEEE.

International Standards Organization (ISO) (1997). ISO 9241-11 (1997). Ergonomische Anforderungen für Bürotätigkeiten mit Bildschirmgeräten Teil 11: Anforderungen an die Gebrauchstauglichkeit – Leitsätze. Deutsche Fassung. EN ISO 9241-11:1997.

Isaksson-Hellman, I., & Lindman, M. (2016). Evaluation of the crash mitigation effect of low-speed automated emergency braking systems based on insurance claims data. *Traffic injury prevention*, *17*(sup1), 42-47.

Jacobs, G., Aeron-Thomas, A., & Astrop, A. (2000). Estimating global road fatalities.

Jamson, A. H., Merat, N., Carsten, O. M., & Lai, F. C. (2013). Behavioural changes in drivers experiencing highly-automated vehicle control in varying traffic conditions. *Transportation research part C: emerging technologies*, *30*, 116-125.

Jasper, H. H. (1958). The ten-twenty electrode system of the International Federation. *Electroencephalogr. Clin. Neurophysiol.*, *10*, 370-375.

Jensen, O., Kaiser, J., & Lachaux, J. P. (2007). Human gamma-frequency oscillations associated with attention and memory. *Trends in neurosciences*, *30*(7), 317-324.

Jorna, P. G. (1992). Spectral analysis of heart rate and psychological state: A review of its validity as a workload index. *Biological psychology*, *34*(2-3), 237-257.

Kahneman, D. (1973). *Attention and effort* (Vol. 1063). Englewood Cliffs, NJ: Prentice-Hall.

Käthner, I., Wriessnegger, S. C., Müller-Putz, G. R., Kübler, A., & Halder, S. (2014). Effects of mental workload and fatigue on the P300, alpha and theta band power during operation of an ERP (P300) brain–computer interface. *Biological psychology*, *102*, 118-129.

Kayikcioglu, T., & Aydemir, O. (2010). A polynomial fitting and k-NN based approach for improving classification of motor imagery BCI data. *Pattern Recognition Letters*, *31*(11), 1207-1215.

Keil, A., Mussweiler, T., & Epstude, K. (2006). Alpha-band activity reflects reduction of mental effort in a comparison task: a source space analysis. *Brain Research*, *1121*(1), 117-127.

Kessel, C. J., & Wickens, C. D. (1982). The transfer of failure-detection skills between monitoring and controlling dynamic systems. *Human Factors*, *24*(1), 49-60.

Kirchner, W. K. (1958). Age differences in short-term retention of rapidly changing information. *Journal of experimental psychology*, *55*(4), 352.

Klimesch, W. (1999). EEG alpha and theta oscillations reflect cognitive and memory performance: a review and analysis. *Brain research reviews*, *29*(2-3), 169-195.

Klimesch, W., Doppelmayr, M., Schwaiger, J., Winkler, T., & Gruber, W. (2000). Theta oscillations and the ERP old/new effect: independent phenomena?. *Clinical Neurophysiology*, *111*(5), 781-793.

Koelega, H. S. (1992). Extraversion and vigilance performance: 30 years of inconsistencies. *Psychological bulletin*, *112*(2), 239.

Koelega, H. S., Verbaten, M. N., Van Leeuwen, T. H., Kenemans, J. L., Kemner, C., & Sjouw, W. (1992). Time effects on event-related brain potentials and vigilance performance. *Biological psychology*, *34*(1), 59-86.

Kohlmorgen, J., Dornhege, G., Braun, M., Blankertz, B., Müller, K. R., Curio, G., ... & Kincses, W. (2007). Improving human performance in a real operating environment through real-time mental workload detection. *Toward Brain-Computer Interfacing*, *409422*.

Kok, A. (2001). On the utility of P3 amplitude as a measure of processing capacity. *Psychophysiology*, *38*(3), 557-577.

Körber, M., Cingel, A., Zimmermann, M., & Bengler, K. (2015). Vigilance decrement and passive fatigue caused by monotony in automated driving. *Procedia Manufacturing*, *3*, 2403-2409.

Koshino, H., Carpenter, P. A., Keller, T. A., & Just, M. A. (2005). Interactions between the dorsal and the ventral pathways in mental rotation: an fMRI study. *Cognitive, Affective, & Behavioral Neuroscience*, *5*(1), 54-66.

Kotchoubey, B. (2006). Event-related potentials, cognition, and behavior: a biological approach. *Neuroscience & Biobehavioral Reviews*, *30*(1), 42-65.

Kothe, C. (2009). Design and Implementation of a Research Brain-Computer Interface. *Berlin Institute of Technology, Berlin.*

Kothe, C. A., & Makeig, S. (2013). BCILAB: a platform for brain–computer interface development. *Journal of neural engineering*, *10*(5), 056014.

Kramer, A. F. (1991). Physiological metrics of mental workload: A review of recent progress. *Multiple-task performance*, 279-328.

Kramer, A. F., Trejo, L. J., & Humphrey, D. (1995). Assessment of mental workload with task-irrelevant auditory probes. *Biological psychology*, *40*(1-2), 83-100.

Krol, L. R., Freytag, S. C., Fleck, M., Gramann, K., & Zander, T. O. (2016, October). A task-independent workload classifier for neuroadaptive technology: Preliminary data. In *2016 IEEE International Conference on Systems, Man, and Cybernetics (SMC)* (pp. 003171-003174). IEEE.

Laurell, H., & Lisper, H. O. (1978). A Validation of Subsidiary Reaction Time Against Detection of Roadside Obstacles during Prolonged Driving∗. *Ergonomics*, *21*(2), 81-88.

Lee, D. H., & Park, K. S. (1990). Multivariate analysis of mental and physical load components in sinus arrhythmia scores. *Ergonomics*, *33*(1), 35-47.

Lei, S. (2011). *Driver mental states monitoring based on brain signals* (Doctoral dissertation, Ph. D. thesis, TU Berlin, Germany).

Lei, S., Toriizuka, T., & Roetting, M. (2017). Driver adaptive task allocation: A field driving study. *Le travail humain*, *80*(1), 93-112.

Lemm, S., Blankertz, B., Curio, G., & Muller, K. R. (2005). Spatio-spectral filters for improving the classification of single trial EEG. *IEEE transactions on biomedical engineering*, *52*(9), 1541-1548.

Llaneras, R. E., Salinger, J., & Green, C. A. (2013). Human factors issues associated with limited ability autonomous driving systems: Drivers' allocation of visual attention to the forward roadway.

Lotte, F., Congedo, M., Lécuyer, A., Lamarche, F., & Arnaldi, B. (2007). A review of classification algorithms for EEG-based brain–computer interfaces. *Journal of neural engineering*, *4*(2), R1.

Louw, T., Kountouriotis, G., Carsten, O., & Merat, N. (2015). Driver inattention during vehicle automation: How does driver engagement affect resumption of control?. In *4th International Conference on Driver Distraction and Inattention (DDI2015), Sydney: proceedings*. ARRB Group.

Makeig, S., Westerfield, M., Jung, T. P., Enghoff, S., Townsend, J., Courchesne, E., & Sejnowski, T. J. (2002). Dynamic brain sources of visual evoked responses. *Science*, *295*(5555), 690-694.

Manzey, D. (1998). Psychophysiologie mentaler Beanspruchung. *Ergebnisse und Anwendungen der Psychophysiologie. Enzyklopädie der Psychologie, 100*, 799-864.

Martens, M. H., & Van Winsum, W. (2000). Measuring distraction: the peripheral detection task. *TNO Human Factors, Soesterberg, Netherlands*.

Mattes, S. (2003). The lane-change-task as a tool for driver distraction evaluation. *Quality of Work and Products in Enterprises of the Future*, *57*, 60.

Maurer, M., Christian Gerdes, J., Lenz, B., & Winner, H. (2016). *Autonomous driving: technical, legal and social aspects*. Springer Nature.

McCallum, W. C., Cooper, R., & Pocock, P. V. (1988). Brain slow potential and ERP changes associated with operator load in a visual tracking task. *Electroencephalography and clinical Neurophysiology*, *69*(5), 453-468.

McFarland, D. J., McCane, L. M., David, S. V., & Wolpaw, J. R. (1997). Spatial filter selection for EEG-based communication. *Electroencephalography and clinical Neurophysiology*, *103*(3), 386-394.

McIntire, L. K., McKinley, R. A., Goodyear, C., & McIntire, J. P. (2014). Detection of vigilance performance using eye blinks. *Applied ergonomics*, *45*(2), 354-362.

Meijman, T. F., & O'Hanlon, J. F. (1984). Workload. An introduction to psychological theories and measurement methods. *Handbook of work and organizational psychology*, *1*, 257-288.

Merat, N., & Jamson, A. H. (2009). How do drivers behave in a highly automated car?.

Merat, N., Jamson, A. H., Lai, F. C., & Carsten, O. (2012). Highly automated driving, secondary task performance, and driver state. *Human factors*, *54*(5), 762-771.

Merat, N., Jamson, A. H., Lai, F. C., Daly, M., & Carsten, O. M. (2014). Transition to manual: Driver behaviour when resuming control from a highly automated vehicle. *Transportation research part F: traffic psychology and behaviour*, *27*, 274-282.

Millan, J. R., & Mouriño, J. (2003). Asynchronous BCI and local neural classifiers: an overview of the adaptive brain interface project. *IEEE transactions on neural systems and rehabilitation engineering*, *11*(2), 159-161.

Miyata, A., Jiang, L., Dahl, R. D., Kitada, C., Kubo, K., Fujino, M., ... & Arimura, A. (1990). Isolation of a neuropeptide corresponding to the N-terminal 27 residues of the pituitary adenylate cyclase activating polypeptide with 38 residues (PACAP38). *Biochemical and biophysical research communications*, *170*(2), 643-648.

Morgan, P., Alford, C., & Parkhurst, G. (2016). Handover issues in autonomous driving: A literature review.

Muckler, F. A., & Seven, S. A. (1992). Selecting performance measures:" Objective" versus" subjective" measurement. *Human factors*, *34*(4), 441-455.

Muller, K. R., Anderson, C. W., & Birch, G. E. (2003). Linear and nonlinear methods for brain-computer interfaces. *IEEE transactions on neural systems and rehabilitation engineering*, *11*(2), 165-169.

Neale, V. L., & Dingus, T. A. (1998). Commentaries in: Human Factors issues for automated highway systems (AHS). *Intelligent Transportation Systems Journal: Technology, Planning, and Operations*, *4*, 111-119.

Niedermeyer, E., & da Silva, F. L. (Eds.). (2005). *Electroencephalography: basic principles, clinical applications, and related fields*. Lippincott Williams & Wilkins.

O'Donnell, R. D., & Eggemeier, F. T. (1986). Workload assessment methodology. Handbook of Perception and Human Performance. Volume 2. Cognitive Processes and Performance. KR Boff, L. Kaufman and JP Thomas.

O'Hanlon, J. F., & Kelley, G. R. (1977). Comparison of performance and physiological changes between drivers who perform well and poorly during prolonged vehicular operation. In *Vigilance* (pp. 87-109). Springer, Boston, MA.

Papadelis, C., Chen, Z., Kourtidou-Papadeli, C., Bamidis, P. D., Chouvarda, I., Bekiaris, E., & Maglaveras, N. (2007). Monitoring sleepiness with on-board electrophysiological recordings for preventing sleep-deprived traffic accidents. *Clinical Neurophysiology*, *118*(9), 1906-1922.

Parasuraman, R., Sheridan, T. B., & Wickens, C. D. (2000). A model for types and levels of human interaction with automation. *IEEE Transactions on systems, man, and cybernetics-Part A: Systems and Humans*, *30*(3), 286-297.

Parasuraman, R., & Rizzo, M. (2007). Introduction to neuroergonomics. In R. Parasuraman & M. Rizzo (Eds.), *Neuroergonomics: The brain at work* (pp. 3–12). New York: Oxford University Press.

Parasuraman, R., Galster, S., Squire, P., Furukawa, H., & Miller, C. (2005). A flexible delegation-type interface enhances system performance in human supervision of multiple robots: Empirical studies with RoboFlag. *IEEE Transactions on systems, man, and cybernetics-part A: Systems and Humans*, *35*(4), 481-493.

Parasuraman, R., Warm, J. S., & See, J. E. (1998). Brain systems of vigilance.

Pataki, K., Schulze-Kissing, D., Mahlke, S., & Thüring, M. (2005). Anwendung von Usability-Maßen zur Nutzeneinschätzung von Fahrerassistenzsystemen. *Beiträge zur Mensch-Maschine-Systemtechnik aus Forschung und Praxis*, 211-228.

Pattyn, N., Neyt, X., Henderickx, D., & Soetens, E. (2008). Psychophysiological investigation of vigilance decrement: Boredom or cognitive fatigue? Physiology & Behavior, 93(1), 369–378.

Pauzié, A. (2008). A method to assess the driver mental workload: The driving activity load index (DALI). *IET Intelligent Transport Systems*, *2*(4), 315-322.

Pellouchoud, E., Smith, M. E., McEvoy, L., & Gevins, A. (1999). Mental effort-related EEG modulation during video-game play: Comparison between juvenile subjects with epilepsy and normal control subjects. *Epilepsia, 40*(Suppl. 4), 38–43.

Penny, W. D., Roberts, S. J., Curran, E. A., & Stokes, M. J. (2000). EEG-based communication: a pattern recognition approach. *IEEE transactions on Rehabilitation Engineering*, *8*(2), 214-215.

Pfurtscheller, G., & Da Silva, F. L. (1999). Event-related EEG/MEG synchronization and desynchronization: basic principles. *Clinical neurophysiology*, *110*(11), 1842-1857.

Pfurtscheller, G., Neuper, C., Schlogl, A., & Lugger, K. (1998). Separability of EEG signals recorded during right and left motor imagery using adaptive autoregressive parameters. *IEEE transactions on Rehabilitation Engineering*, *6*(3), 316-325.

Philip, P., Sagaspe, P., Taillard, J., Moore, N., Guilleminault, C., Sanchez-Ortuno, M., ... & Bioulac, B. (2003). Fatigue, sleep restriction, and performance in automobile drivers: a controlled study in a natural environment. *Sleep*, *26*(3), 277-280.

Picton, T. W., Bentin, S., Berg, P., Donchin, E., Hillyard, S. A., Johnson Jr, R., ... & Taylor, M. J. (2000). Guidelines for using human event-related potentials to study cognition: Recording standards and publication criteria. *Psychophysiology*, *37*(2), 127-152.

Pomerleau, D. A. (1993). Knowledge-based training of artificial neural networks for autonomous robot driving. In *Robot learning* (pp. 19-43). Springer, Boston, MA.

Pratt, N., Willoughby, A., & Swick, D. (2011). Effects of working memory load on visual selective attention: behavioral and electrophysiological evidence. *Frontiers in human neuroscience*, *5*, 57.

Prinzel, L. J., DeVries, H., Freeman, F. G., & Mikulka, P. (2001). Examination of automation-induced complacency and individual difference variates. National Aeronautics and Space Administration, Langley Research Center, Hampton, Virginia. Retrieved from http:// www.cs.odu.edu/~mln/ltrs-pdfs/NASA-2001-tm211413.pdf

Prinzel, L. J., Freeman, F. G., Scerbo, M. W., Mikulka, P. J., & Pope, A. T. (2000). A closed-loop system for examining psychophysiological measures for adaptive task allocation. *The International journal of aviation psychology*, *10*(4), 393-410.

Qu, W., Ge, Y., Xiong, Y., Carciofo, R., Zhao, W., & Zhang, K. (2015). The relationship between mind wandering and dangerous driving behavior among Chinese drivers. *Safety Science*, *78*, 41-48.

Radlmayr, J., Gold, C., Lorenz, L., Farid, M., & Bengler, K. (2014, September). How traffic situations and non-driving related tasks affect the take-over quality in highly automated driving. In *Proceedings of the human factors and ergonomics society annual meeting* (Vol. 58, No. 1, pp. 2063-2067). Sage CA: Los Angeles, CA: Sage Publications.

Raghavachari, S., Kahana, M. J., Rizzuto, D. S., Caplan, J. B., Kirschen, M. P., Bourgeois, B., ... & Lisman, J. E. (2001). Gating of human theta oscillations by a working memory task. *Journal of Neuroscience*, *21*(9), 3175-3183.

Ramoser, H., Muller-Gerking, J., & Pfurtscheller, G. (2000). Optimal spatial filtering of single trial EEG during imagined hand movement. *IEEE transactions on rehabilitation engineering*, *8*(4), 441-446.

Rauch, N., Kaussner, A., Kruger, H. P., Boverie, S., & Flemisch, F. (2009). The importance of driver state assessment within highly automated vehicles. In *16th ITS World Congress and Exhibition on Intelligent Transport Systems and ServicesITS AmericaERTICOITS Japan*.

Rechtschaffen, A. (1968). A manual for standardized terminology, techniques and scoring system for sleep stages in human subjects. *Brain information service*.

Redick, T. S., & Lindsey, D. R. (2013). Complex span and n-back measures of working memory: A meta-analysis. *Psychonomic bulletin & review*, *20*(6), 1102-1113.

Reid, G. B., Shingledecker, C. A., & Eggemeier, F. T. (1981, October). Application of conjoint measurement to workload scale development. In *Proceedings of the Human Factors Society Annual Meeting* (Vol. 25, No. 1, pp. 522-526). Sage CA: Los Angeles, CA: SAGE Publications.

Rhede, J. G. (2017). *Konzeption und Evaluation einer hochintegrativen Anzeige für Fahrassistenzsysteme im Pkw in einer handlungsorientierten Warnstrategie*. Technische Universitaet Berlin (Germany).

Robertson, I. H., Manly, T., Andrade, J., Baddeley, B. T., & Yiend, J. (1997). Oops!': performance correlates of everyday attentional failures in traumatic brain injured and normal subjects. *Neuropsychologia*, *35*(6), 747-758.

Rösler, F., Heil, M., & Röder, B. (1997). Slow negative brain potentials as reflections of specific modular resources of cognition. *Biological psychology*, *45*(1-3), 109-141.

Rötting, M. (2001). *Parametersystematik der Augen-und Blickbewegungen für arbeitswissenschaftliche Untersuchungen*. Shaker.

Rouse, W. B., Edwards, S. L., & Hammer, J. M. (1993). Modeling the dynamics of mental workload and human performance in complex systems. *IEEE transactions on systems, man, and cybernetics*, *23*(6), 1662-1671.

Rubio, S., Díaz, E., Martín, J., & Puente, J. M. (2004). Evaluation of subjective mental workload: A comparison of SWAT, NASA-TLX, and workload profile methods. *Applied Psychology*, *53*(1), 61-86.

Ruchkin, D. S., Johnson Jr, R., Canoune, H., & Ritter, W. (1990). Short-term memory storage and retention: An event-related brain potential study. *Electroencephalography and clinical Neurophysiology*, *76*(5), 419-439.

Ruff, S. (2017). *Driver Cognitive Workload: A Comprehensive Multi-measure Approach*. Technische Universitaet Berlin (Germany).

Saffarian, M., De Winter, J. C., & Happee, R. (2012, September). Automated driving: human-factors issues and design solutions. In *Proceedings of the human factors and ergonomics society annual meeting* (Vol. 56, No. 1, pp. 2296-2300). Sage CA: Los Angeles, CA: Sage Publications.

Scerbo, M. W., Freeman, F. G., & Mikulka, P. J. (2003). A brain-based system for adaptive

automation. *Theoretical Issues in Ergonomics Science*, *4*(1-2), 200-219.

Schlögl, A. (2000). *The electroencephalogram and the adaptive autoregressive model: theory and applications*. Aachen: Shaker.

Schmidt, E. A., Kincses, W. E., Scharuf, M., Haufe, S., Schubert, R., & Curio, G. (2007). Assessing drivers' vigilance state during monotonous driving.

Schmidt, E. A., Schrauf, M., Simon, M., Fritzsche, M., Buchner, A., & Kincses, W. E. (2009). Drivers' misjudgement of vigilance state during prolonged monotonous daytime driving. *Accident Analysis & Prevention*, *41*(5), 1087-1093.

Schubert, M. N. (2019). Regulating the Use of Automated Vehicles (SAE Levels 3 to 5) in Germany and the UK. *RAW, 3*.

Shladover, S. E., Nowakowski, C., Lu, X. Y., & Ferlis, R. (2015). Cooperative adaptive cruise control: Definitions and operating concepts. *Transportation Research Record*, *2489*(1), 145-152.

Sibi, S., Ayaz, H., Kuhns, D. P., Sirkin, D. M., & Ju, W. (2016, June). Monitoring driver cognitive load using functional near infrared spectroscopy in partially autonomous cars. In *2016 IEEE Intelligent Vehicles Symposium (IV)* (pp. 419-425). IEEE.

Smith, M. E., Gevins, A., Brown, H., Karnik, A., & Du, R. (2001). Monitoring task loading with multivariate EEG measures during complex forms of human-computer interaction. *Human Factors*, *43*(3), 366-380.

So, W. K., Wong, S. W., Mak, J. N., & Chan, R. H. (2017). An evaluation of mental workload with frontal EEG. *PloS one*, *12*(4), e0174949.

Statistisches Bundesamt. (2017). *Verkehr: Verkehrsunfäl le.* Zugriff am 30.10.2017 auf https://www.destatis.de/DE/Publikationen/Thematisch/TransportVerkehr/Verkehrsunfaelle/VerkehrsunfaellleJ2080700167004 .pdf?__blob=publicationFile

Sterman, M. B., Kaiser, D. A., Mann, C. A., Suyenobu, B. Y., Beyma, D. C., & Francis, J. R. (1993, October). Application of quantitative EEG analysis to workload assessment in an advanced aircraft simulator. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 37, No. 1, pp. 118-121). Sage CA: Los Angeles, CA: SAGE Publications.

Syam, S. H. F., Lakany, H., Ahmad, R. B., & Conway, B. A. (2017, December). Comparing common

average referencing to laplacian referencing in detecting imagination and intention of movement for brain computer interface. In *MATEC Web of Conferences* (Vol. 140).

Tallon-Baudry, C., Bertrand, O., Hénaff, M. A., Isnard, J., & Fischer, C. (2005). Attention modulates gamma-band oscillations differently in the human lateral occipital cortex and fusiform gyrus. *Cerebral cortex*, *15*(5), 654-662.

Teplan, M. (2002). Fundamentals of EEG measurement. *Measurement science review*, *2*(2), 1-11.

Thiffault, P., & Bergeron, J. (2003). Monotony of road environment and driver fatigue: a simulator study. *Accident Analysis & Prevention*, *35*(3), 381-391.

Thomas, H. B. G. (1963). Communication theory and the constellation hypothesis of calculation. *Quarterly Journal of Experimental Psychology*, *15*(3), 173-191.

Thorpe, C., Herbert, M., Kanade, T., & Shafer, S. (1991). Toward autonomous driving: the cmu navlab. i. perception. *IEEE expert*, *6*(4), 31-42.

Turner, M. L., & Engle, R. W. (1989). Is working memory capacity task dependent?. *Journal of memory and language*, *28*(2), 127-154.

Ullsperger, M., & Von Cramon, D. Y. (2001). Subprocesses of performance monitoring: a dissociation of error processing and response competition revealed by event-related fMRI and ERPs. *Neuroimage*, *14*(6), 1387-1401.

Unde, S. A., & Shriram, R. (2014, April). Coherence analysis of EEG signal using power spectral density. In *Proceedings of the 2014 Fourth International Conference on Communication Systems and Network Technologies* (pp. 871-874).

Unema, P., Rötting, M., Sepher-Willeberg, M., Strümpfel, U. & Kopp, U. (1988). Der NASA Task Load Index: Erste Ergebnisse mit der deutschen Fassung. In Gesellschaft für Arbeitswissenschaft e.V. (Hrsg.). *Jahresdokumentation 1988 der Gesellschaft für Arbeitswissenschaft e.V. – Bericht zum 34. Arbeitswissenschaftlichen Kongress an der RWTH Aachen*. Köln: O. Schmidt.

Van Der Laan, J. D., Heino, A., & De Waard, D. (1997). A simple procedure for the assessment of acceptance of advanced transport telematics. *Transportation Research Part C: Emerging Technologies*, *5*(1), 1-10.

Verwey, W. B. (1990). *Adaptable driver-car interfacing and mental workload: a review of the literature* (No. IZF-1990-B-3). Institute for Perception RVO-TNO Soesterberg (Netherlands).

Vidulich, M. A., & Tsang, P. S. (2012). Mental workload and situation awareness.(chap. 8), and Salvendy, G.(Ed.). Handbook of Human Factors and Ergonomics.

Vigário, R., Sarela, J., Jousmiki, V., Hamalainen, M., & Oja, E. (2000). Independent component approach to the analysis of EEG and MEG recordings. *IEEE transactions on biomedical engineering*, *47*(5), 589-593.

Walter, C., Rosenstiel, W., Bogdan, M., Gerjets, P., & Spüler, M. (2017). Online eeg-based workload adaptation of an arithmetic learning environment. *Frontiers in human neuroscience*, *11*, 286.

Warm, J. S., Matthews, G., & Finomore Jr, V. S. (2008). Vigilance, workload, and stress. *Performance under stress*, 115-41.

Welke, S. (2012). *Lenkmanöverprädiktion basierend auf einer Analyse der hirnelektrischen Aktivität des Fahrers* (Doctoral dissertation, Universitätsbibliothek der Technischen Universität Berlin).

Wickens, C. D. (2002). Multiple resources and performance prediction. *Theoretical issues in ergonomics science*, *3*(2), 159-177.

Wickens, C. D. (2008). Multiple resources and mental workload. *Human Factors,* 50 (3), 449-455.

Wierwille, W. W., & Eggemeier, F. T. (1993). Recommendations for mental workload measurement in a test and evaluation environment. *Human factors*, *35*(2), 263-281.

Williams, H. L., Lubin, A., & Goodnow, J. J. (1960). Other: Impaired Performance with Acute Sleep Loss. Psychological Monographs: General and Applied. *Nursing Research*, *9*(4), 234.

Wilson, G. F., & Russell, C. A. (2004). Psychophysiologically determined adaptive aiding in a simulated UCAV task. *Human performance, situation awareness, and automation: Current research and trends*, 200-204.

Wyatt, H. J., & Musselman, J. F. (1981). Pupillary light reflex in humans: evidence for an unbalanced pathway from nasal retina, and for signal cancellation in brainstem. *Vision research*, *21*(4), 513-525.

Xiao, Y. M., Wang, Z. M., Wang, M. Z., & Lan, Y. J. (2005). The appraisal of reliability and validity of subjective workload assessment technique and NASA-task load index. *Zhonghua lao dong wei sheng zhi ye*

bing za zhi= Zhonghua laodong weisheng zhiyebing zazhi= Chinese journal of industrial hygiene and occupational diseases, *23*(3), 178-181.

Yoon, S. H., Kim, Y. W., & Ji, Y. G. (2019). The effects of takeover request modalities on highly automated car control transitions. *Accident Analysis & Prevention*, *123*, 150-158.

Young, M. S., & Stanton, N. A. (1997). Automotive automation: Investigating the impact on drivers' mental workload.

Young, M. S., & Stanton, N. A. (2002). Attention and automation: new perspectives on mental underload and performance. *Theoretical issues in ergonomics science*, *3*(2), 178-194.

Zander, T. O. (2012). Utilizing brain-computer interfaces for human-machine systems dissertation. *TU Berlin*.

Zander, T. O., Lehne, M., Ihme, K., Jatzev, S., Correia, J., Kothe, C., ... & Nijboer, F. (2011). A dry EEG-system for scientific research and brain–computer interfaces. *Frontiers in neuroscience*, *5*, 53.

Zander, T. O., Andreessen, L. M., Berg, A., Bleuel, M., Pawlitzki, J., Zawallich, L., ... & Gramann, K. (2017). Evaluation of a dry EEG system for application of passive brain-computer interfaces in autonomous driving. *Frontiers in human neuroscience*, *11*, 78.

Zeeb, K., Buchner, A., & Schrauf, M. (2015). What determines the take-over time? An integrated model approach of driver take-over after automated driving. *Accident analysis & prevention*, *78*, 212-221.

List of figures

Figure 2.1 Levels of vehicle automation, which are modified based on definitions from SAE (Driving, 2014) and BASt (Gasser et al., 2012).....20

Figure 2.2 An electric cart developed by Stanford Artificial Intelligence Laboratory in 1960s (left, Credit: @ Mark Richards), and a prototype of autonomous driving developed by Japan's Tsukuba Mechanical Engineering Laboratory in 1977 (right) 22

Figure 2.3 An Audi A8 launched in 2018 with Traffic Jam Assistant System.....24

Figure 5.1 Illustration of human brain and the structure of the origin of electrical activity within the brain (source: Welke (2012), which modified based on Hämäläinen et al. (1993) and Vigário et al. (2000)).....44

Figure 5.2 Placement of EEG electrodes according to the 10-20 system (left) and electrode positions on the scalp in a top view (right) (source: Welke (2012), according to Jasper's (1958) 10-20 system) 46

Figure 5.3 EEG frequency components delta, theta, alpha, beta and gamma.....47

Figure 5.4 Typical Event-Related Potential waveform.....48

Figure 5.5 Typical changes in spectral power of EEG in the 4-14 Hz range at frontal (Fz) and parietal (Pz) midline electrodes under high and low workload conditions (Data from Gevins & Smith, 2000).....50

Figure 6.1 Online processing procedures (Duda et al., 2000) 53

Figure 6.2 A typical pair of CSP patterns in the motor imagery study. The pattern on the left hand side shows weights that are optimized for imagined left hand movements. The other pattern is optimized for right hand movements. Red indicates positive weights, while negative weights are in blue. (Zander, 2012) 58

Figure 6.3 Architecture of the Filter Bank Common Spatial Pattern (FBCSP) approach (Ang et al., 2008) 59

Figure 6.4 A hyperplane separating two classes. The crosses are one class, and the circles are the other class......62

Figure 7.1 Average NASA-TLX ratings across all participants in each task.....68

Figure 7.2 General procedure of Experiment 1 ..... *70*

Figure 7.3 Average NASA-TLX ratings across all participants in each task *72*

Figure 7.4 Averaged classifier outputs across participants *75*

Figure 7.5 First CSP pattern pairs of two frequency bands: theta band (4-7 Hz) in the first row and alpha band (8-13 Hz) in the second row (blue means negative, yellow means positive) *77*

Figure 8.1 NASA-TLX ratings of different tasks *85*

Figure 8.2 SEA-Scale ratings of different tasks *86*

Figure 8.3 Detection rate in auditory oddball task *87*

Figure 9.1 General procedure of Experiment 2 ..... *91*

Figure 9.2 Driving simulator – A real driving vehicle *92*

Figure 9.3 NeuroScan wet electrode cap *92*

Figure 9.4 Higher-resolution international 10-20 system *93*

Figure 9.5 A participant wearing EEG system seated in the simulator vehicle *94*

Figure 9.6 Subjective ratings of NASA-TLX *96*

Figure 9.7 Subjective ratings of SEA-Scale *97*

Figure 9.8 Averaged classifier outputs across participants *103*

Figure 10.1 Illustration of procedure of dual tasks *110*

Figure 10.2 Procedure of the oddball task in one dual task block. The sequence of targets and distractors in this figure is merely an example. The inter-stimulus interval (ISI) was randomly chosen from 5 to 10 seconds, so that participants had to react every 30 seconds on average. *111*

Figure 10.3 Grand average of targets and distractors at electrode positions Fz, FCz, Cz and Pz across all participants and all tasks *113*

Figure 10.4 Grand average of hits, misses and correct rejections at electrode positions Fz, Cz and Pz, across all participants and all tasks.....115

Figure 10.5 Average rates of hits, misses, false alarms, correct rejections and accuracies of all participants under different task conditions.....118

Figure 10.6 Miss rates under seven task conditions. Among them, tasks of different workload levels are also presented in different colors. Paired t-tests were calculated and significances were marked.....119

Figure 10.7 Reaction time of oddball task under different task conditions 120

Figure 11.1 Illustration of independent variables in the experimental design 125

Figure 11.2 General procedure of experiment 3.....126

Figure 11.3 Procedure of one online feedback block and one trial in the block 127

Figure 11.4 CARLA simulator environment.....128

Figure 11.5 Disengagement button on the steering wheel for take-over 128

Figure 11.6 SEA-scale ratings of four tasks 130

Figure 11.7 Take-over time of high and low workload tasks.....133

Figure 11.8 Take-over time of high workload tasks in with & no feedback sessions 134

Figure 11.9 Subjective ratings on the appearance of feedback warnings 135

List of tables

Table 7.1 Average ratings of NASA-TLX across all participants and comparison results between two difficulty levels in each task *68*

Table 7.2 Average ratings of NASA-TLX across all participants and comparison results between two difficulty levels in each task *71*

Table 7.3 Classification accuracies of High vs. Low workload conditions *73*

Table 7.4 Mean classifier outputs of each tasks in High and Low workload conditions *74*

Table 7.5 Results of permutation tests on outputs of High vs. Low workload conditions of each participant *76*

Table 8.1 p-values of t-test of NASA-TLX rating scores *85*

Table 8.2 p-values of t-test of SEA-Scale rating scores *86*

Table 9.1 Subjective reporting scores of NASA-TLX & SEA-Scale *95*

Table 9.2 p-values of t-tests of NASA-TLX rating scores *98*

Table 9.3 p-values of t-tests of SEA scale ratings *98*

Table 9.4 Classification accuracies of Low vs. High condition tasks *99*

Table 9.5 Classification accuracies of Low vs. Medium conditions *100*

Table 9.6 Classification accuracies of Medium vs. High conditions *101*

Table 9.7 Classification accuracy after applying task-independent classifier (%) *102*

Table 9.8 Averaged classifier outputs across participants *103*

Table 9.9 Results of permutation tests on outputs of High vs. Low workload conditions *104*

Table 10.1 Mean classification accuracies of targets vs. distractors across all participants in different task blocks (%) *114*

Table 10.2 Mean classification accuracies of hits vs. correct rejections across all participants in different task blocks (%) *116*

Table 10.3 Rates of hits, misses, false alarms, correct rejections, and accuracies under seven different task conditions and average across all these task conditions (%) *117*

Table 10.4 Reaction time of oddball task under different task conditions and their mean *119*

Table 11.1 SEA-scale ratings of four tasks *130*

Table 11.2 Classification accuracy between high and low workload conditions (%) *131*

Table 11.3 Feedback appearance rate *132*

Table 11.4 Take-over time (s) of high and low workload tasks and their comparison results *133*

Table 11.5 Take-over time (s) of high workload tasks in with & no feedback sessions and their comparison results *134*

Table 11.6 Subjective ratings on the appearance of feedback warnings *135*

List of abbreviations

ABS: Anti-lock Brake System

ACC: Adaptive Cruise Control

ADAS: Advanced Driver Assistance System

AEB: Autonomous Emergency Braking

ANN: Artificial Neural Networks

ANS: Autonomic Nervous System

AR: Autoregressive

BASt: Bundesanstalt für Straßenwesen

BCI: Brain-Computer-Interface

CAR: Common Average Reference

CNS: Central Nervous System

CSP: Common Spatial Patterns

CSSP: Common Spatio-Spectral Pattern

CSSSP: Common Sparse Spectral Pattern

DARPA: Defense Advanced Research Projects Agency

ECG: Electrocardiogram

EDA: Electrodermal Activity

EEG: Electroencephalogram

EOG: Electrooculogram

ERP: Event-Related Potentials

FBCSP: Filter Bank Common Spatial Pattern

FCW: Forward Collision Warning

FFT: Fast Fourier Transform

FIR: Finite Impulse Response

fMRI: Functional Magnetic Resonance Imaging

HAD: Highly Automated Driving

HR: Heart Rate

HRV: Heart Rate Variability

ICA: Independent Component Analysis

IIR: Infinite Impulse Response

IVT: In-Vehicle Technology

LDA: Linear Discriminant Analysis

MATB: Multi-Attribute Task Battery

MLP: Multilayer Perceptron

NASA-TLX: NASA Task Load Index

OEM: Original Equipment Manufacturer

OOTL: Out of the Loop

PCA: Principal Component Analysis

PERCLOS: Percentage of Eye Closure

PNS: Peripheral Nervous System

PSD: Power Spectral Density

SAE: Society of Automotive Engineers

SEA: Skala zur Erfassung von subjektiv erlebter Anstrengung

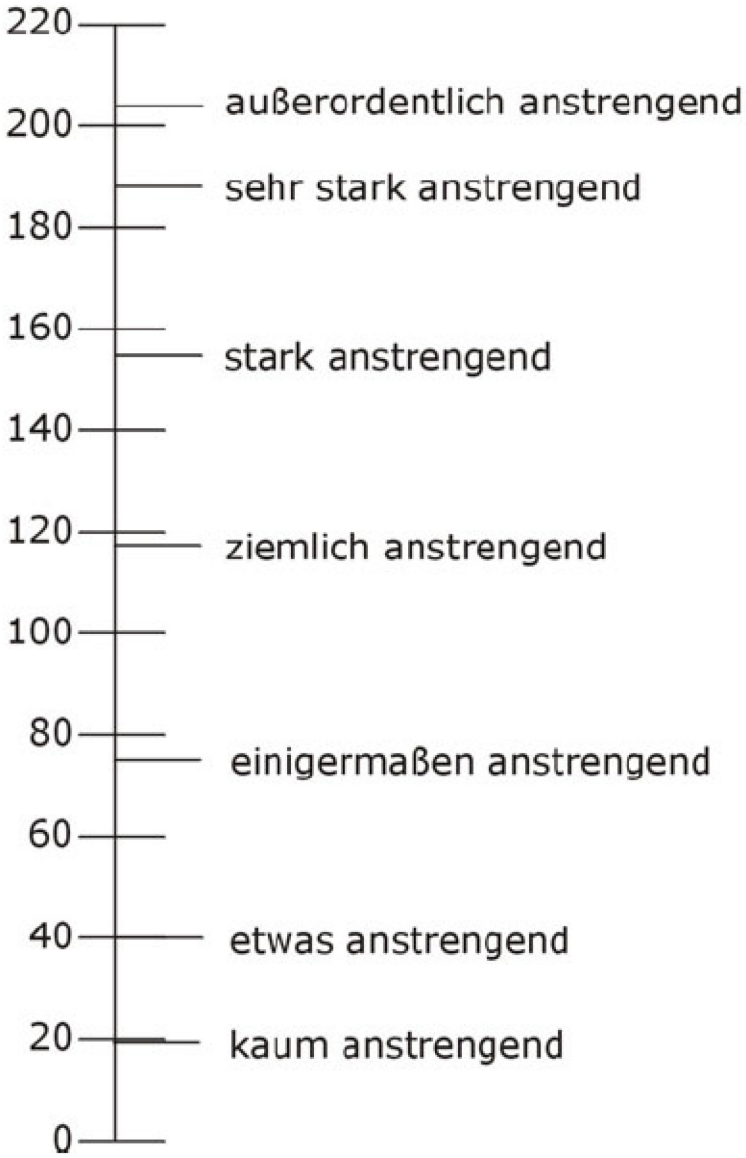
SVM: Support Vector Machine

Appendix

A.1 SEA-Scale

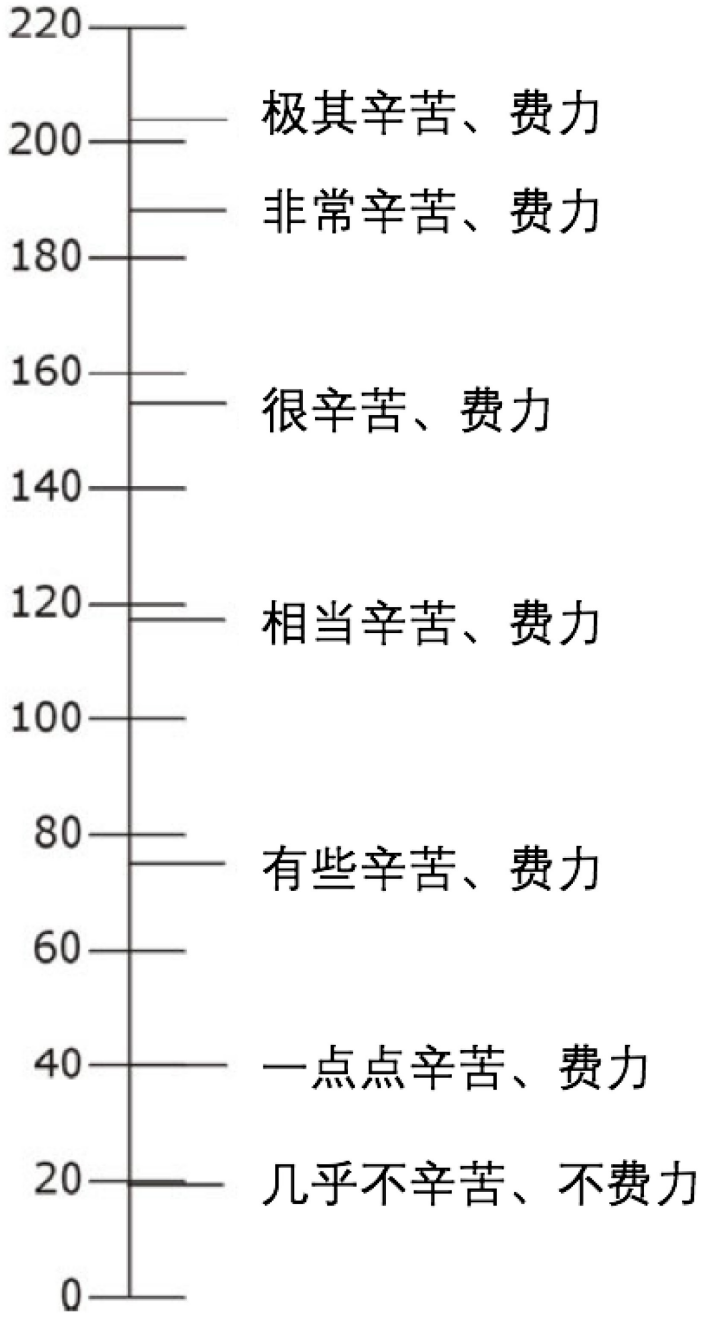
A.1.1 German version

Bitte kreuzen Sie auf der folgenden Skala Ihre Gesamtbewertung für die gerade absolvierte Versuchsbedingung an.



A.1.2 Chinese version translated from German

请评价您在完成刚才的任务时的辛苦、疲劳程度。请在下面刻度符合的位置处打叉。



A.2 NASA-TLX

A.2.1 German version

Geistige Anforderungen

Wie hoch waren die geistigen Anforderungen der Aufgabe?

□ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □

sehr sehr

niedrig hoch

Körperliche Anforderungen

Wie hoch waren die körperlichen Anforderungen der Aufgabe?

□ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □

sehr sehr

niedrig hoch

Zeitliche Anforderungen

Wie hoch war das Tempo, mit dem die einzelnen Arbeitsschritte der Aufgabe aufeinander folgten?

□ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □

sehr sehr

niedrig hoch

Leistung

Wie erfolgreich haben Sie die geforderte Aufgabe Ihrer Meinung nach durchgeführt?

□ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □

sehr sehr

gut schlecht

Anstrengung

Wie sehr mussten Sie sich anstrengen, um Ihre Leistung zu erreichen?

□ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □

sehr sehr

niedrig hoch

Frustration

Wie verunsichert, entmutigt, gereizt und verärgert waren Sie?

□ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □

sehr sehr

niedrig hoch

A.2.2 Chinese version

脑力需求

任务的脑力需求怎么样?

□ □

非常低 非常高

身体负担

任务的体力需求怎么样?

□ □

非常低 非常高

时间需求

任务的进度让你感到匆忙吗?

□ □

非常低 非常高

任务绩效

你在这项任务上做得有多成功?

□ □

完美　　失败

努力程度

为了在这项任务上达到你的水平，你需要付出多大努力？

□ □

非常低　　非常高

挫败感

你有感到任何不安全、气馁、压力或烦恼吗？

□ □

非常低　　非常高