The alpha band as an electrophysiological indicator for internalized attention and high mental workload in real traffic driving.

Inaugural-Dissertation

zur

Erlangung des Doktorgrades der

Mathematisch-Naturwissenschaftlichen Fakultät

der Heinrich-Heine Universität Düsseldorf

vorgelegt von

Konrad Hagemann

aus Ebersberg

1. April 2008

Aus dem Institut für Experimentelle Psychologie

der Heinrich-Heine Universität Düsseldorf

Gedruckt mit der Genehmigung der

Mathematisch-Naturwissenschaftlichen Fakultät der

Heinrich-Heine-Universität Düsseldorf

Referent: Prof. Dr. Axel Buchner

Koreferent: PD. Dr. Michael Schrauf

Tag der mündlichen Prüfung: 09. Mai 2008

Table of Contents

Tables

Figures

Preface

First and foremost, I would like to thank Dr. Wilhelm Kincses and Dr. Michael Schrauf at DaimlerChrysler, Research & Technology in Böblingen and Prof. Dr. Axel Buchner at the Heinrich-Heine University Düsseldorf for their competent and farsighted supervision of this doctoral thesis. They always had an open ear for discussions on the topic of this thesis that were invaluable for the successful completion of this thesis.

This work would not have been possible without the support from my team colleagues at DaimlerChrylser who, above all, spent numerous hours on the test tracks with me and I cordially thank them all! Especially, Siegfried Rothe and Sven Willmann at DaimlerChrysler who played an important key role in the realization of the complex field experiments. Moreover, I am grateful to Eike Schmidt for his detailed comments on earlier versions of the manuscript and his untiring willingness for discussions on the topic. I also thank Rico Saupe who helped me out countless times when I had reached just another dead end in my Matlab scripts. I am also indebted to Dr. Andreas Bruns who was always open to share his expertise in signal analysis with me.

In addition, I thank Prof. Dr. Christoph Braun for his expert comments on the EEG data results and Dr. Kevin Dixon for his evaluation of the CAN Bus data. Moreover, I am indebted to Caroline Schaller, Gunnar Regenbrecht, Jasmin Steinwender, Liane Siebenhaar, Sebastian Belz, and many others for their active support in running the experiments. I would also like to thank Ralf Höger and Norman Cleesattel Sr. who reviewed earlier versions of the manuscript. Last but not least, I would like to thank my family, Christine Hagemann and Wilfried Hasse, for their moral and financial support as well as their understanding and encouragement, especially during the finishing time period of this work.

Deutsche Zusammenfassung

Viele vorhergehende Untersuchungen haben gezeigt, dass Autofahrer, die neben ihrer Fahraufgabe unter dem Einfluss einer Zusatzaufgabe stehen, dadurch sehr schnell und zu einem hohen Grad vom Verkehrsgeschehen abgelenkt sein können, was zu einem erhöhten Unfallrisiko führen kann (z.B. McEvoy et al., 2005). Obwohl das EEG aufgrund seiner hohen zeitlichen Auflösung ein großes Potential als Messinstrument zur Untersuchung dieser Fragestellung bietet, wurde es in diesem Kontext bisher nur unzureichend eingesetzt. In einer Serie von drei Experimenten, die im Labor und in zwei realen Fahrversuchen durchgeführt wurden, konnte diese Forschungslücke nun mit der vorliegenden Arbeit weiter geschlossen werden. Über alle drei Experimente hinweg zeigte sich eine Erhöhung im EEG Alphaband in Folge mentaler Beanspruchung, die durch eine Zuhöraufgabe ausgelöst wurde. Diese Ergebnisse zeigten sich im Labor, sowohl wenn die belastenden Aufgaben allein, als auch wenn diese parallel zu einer simulierten Spurwechselaufgabe durchgeführt wurden. Dieselben Alphaband Ergebnisse zeigten sich auch im realen Straßenverkehr unter Verwendung einer maskierten Zuhöraufgabe ohne Wortdetektion, jedoch nicht bei einer zusätzlich eingeführten Kopfrechenaufgabe. Das dritte Fahrexperiment repliziert die EEG-Alphabandeffekte über zwei Messzeitpunkte unter Verwendung derselben Aufgabe wie in Experiment 1. Die durchgehende Übereinstimmung der Ergebnisse spricht für die hohe Reliabilität und Generalisierbarkeit der Effekte. Die beobachtete Alphabanderhöhung unter mentaler Beanspruchung steht im Einklang mit den Ergebnissen vorhergehender EEG-Laborstudien zu internalisierter Aufmerksamkeit (z.B. Cooper, Croft, Dominey, Burgess, & Gruzelier, 2003; Jensen, Gelfand, Kounios, & Lisman, 2002; Ray & Cole, 1985b), welche zeigen konnten, dass Alphaoszillationen mit der Inhibition aufgabenirrelevanter kortikaler Areale in Zusammenhang stehen. Beim Transfer

dieser Erkenntnisse auf den Kontext des Autofahrens, zeigt sich, dass diese Ergebnisse neue Erklärungsansätze für die zugrundeliegenden neuronalen Mechanismen für Phänomene wie „endogene Ablenkung“ (Recarte & Nunes, 2003) bzw. „unaufmerksame Blindheit“ (Strayer & Drews, 2007) liefern, die in der verkehrspsychologischen Literatur aktuell umfangreich diskutiert werden.

Die zusätzliche Analyse von EKPs, die durch die Stimuli einer Drittaufgabe in Experiment 2 hervorgerufen wurden, ergab ein workloadspezifisches Ergebnismuster, das zwischen Zuhöraufgabe und Kopfrechenaufgabe unterschiedlich war. Für Abschnitte mit gleichzeitiger Kopfrechenaufgabe wurde eine Erhöhung der aufmerksamkeitssensitiven N1 Komponente und eine Reduktion der P3 Komponente demonstriert. EKPs in Abschnitten mit paralleler Zuhöraufgabe zeigten bei der N1 eine Latenzverzögerung von ca. 50 ms und eine damit verbundene Amplitudenreduktion, die offensichtlich die Verarbeitung auditorischer Stimuli widerspiegelt, die auf demselben sensorischen Kanal konkurrieren. Diese Ergebnisse zeigen, dass EKPs im vorliegenden Fall sensitiver als das EEG Alphaband gegenüber der unterschiedlichen Beanspruchungsmanipulation durch die Zweitaufgaben waren. Allerdings mussten an dieser Stelle auch gewisse Fehler im Design der Kopfrechenaufgabe eingeräumt werden, die unter Umständen die Nulleffekte im Alphaband erklären könnten. Dennoch bestätigen die unterschiedlichen Ergebnisse in den EKPs, dass die beiden Aufgaben unterschiedliche Arten von Beanspruchung ansprechen. Modellvorstellungen, die eine Interferenz von eingehenden Signalen in derselben Modalität vorhersagen, werden dabei ebenso bestätigt (z.B. Wickens, 1984).

Im Sinne eines Mehrebenenansatzes (Manzey, 1998) wurden in allen drei Experimenten neben dem EEG auch Verhaltensdaten und EKG Daten erhoben. In Experiment 1 war eine Wortdetektionsaufgabe in die Zuhöraufgabe integriert, die mit einer einfachen Tondetektionsaufgabe verglichen wurde, während in

Experiment 2 die Antworten zur Drittaufgabe als zusätzliches Beanspruchungsmaß ausgewertet werden konnten. Eine höhere Fehlerquote und langsamere Reaktionszeiten unter hoher Beanspruchung bestätigen die erfolgreiche Beanspruchungsmanipulation durch die belastenden Zweitaufgaben während der Fahrt. Ebenso zeigt sich in der Spurwechselaufgabe im Laborexperiment unter hoher Beanspruchung eine höhere Abweichung der Fahrer von der Ideallinie, was ebenfalls bereits bekannten Forschungsergebnissen entspricht (Mattes, 2003).

Die EKG Daten zeigen im Feldexperiment (Experiment 2) unter hoher Beanspruchung eine Erhöhung der Herzrate und ein Absenken der Herzratenvariabilität, was ein in der Fachliteratur vielfach zu beobachtendes Ergebnismuster darstellt (siehe z.B. zusammenfassend Ribback, 2003). Die Ergebnisse aus Experiment 1 sind hingegen uneindeutig. So zeigt sich zwar unter der höheren Beanspruchung eine ebensolche Absenkung der Herzratenvariabilität, aber auch ein Absenken der Herzrate. Auch in Experiment 3 ist das EKG nicht sensitiv gegenüber der Beanspruchungsmanipulation aufgrund der Zweitaufgabe. Allgemein wird zu Bedenken gegeben, dass in den vorliegenden Untersuchungen keine Kontrolle der Atmung erhoben wurde, was zu einer Konfundierung der Daten geführt haben könnte. Die EKG Daten sind daher nur eingeschränkt interpretierbar. Zukünftige Untersuchungen sollten eine solche Kontrollmessung unbedingt mit einschließen.

Ferner ergab die Gegenüberstellung verschiedener Baselinebedingungen (während der Fahrt und in Ruhe mit geschlossenen Augen) mit den Workloadbedingungen Mehrfachzusammenhänge zwischen verschiedenen kognitiven Zuständen und Veränderungen in der Alphapower. In der vorliegenden Arbeit wurde auf Basis dieser Erkenntnisse eine aufgabenspezifische und individuelle Adjustierung des EEG Alphabandes durchgeführt. Ähnliche Prozeduren sind aus vorhergehender Laborforschung bekannt (z.B. Klimesch, Doppelmayr, Russegger, & Pachinger, 1996). Im anwendungsbezogenen Forschungskontext erscheint dieses Vorgehen generell empfehlenswert, um ein optimales Signal-Rausch-Verhältnis zu erzielen. Allerdings wurde auch deutlich, dass weiterer Forschungsbedarf notwendig ist, um Theorien zu entwickeln, die eine bessere Abgrenzung von workloadspezifischem und müdigkeitsbedingtem Alpha erlauben.

Abschließend lässt sich sagen, dass mit der vorliegenden Arbeit eine Brücke geschlagen wurde zwischen reiner grundlagenbasierten Laborforschung zum Thema Aufmerksamkeit auf der einen und der praxisnahen Verkehrspsychologie auf der anderen Seite. Vielfach wurde kritisiert, dass Laboraufgaben nur eine unzureichende Involviertheit der Probanden in die Aufgabe zur Folge haben (z.B. Hole, 2007). Die Stabilität der Ergebnisse spricht für den weitergehenden Einsatz der EEG Messmethode im Feldversuch im realen Straßenverkehr. Auch hat die Frequenzband-Messung in der Regel den Vorteil, dass sie anders als die ereigniskorrelierte P300 Messung auch ohne eine künstliche Zusatzaufgabe auskommen kann. Für die Forschung und Entwicklung in der Fahrer-Fahrzeug-Interaktion kann das EEG für eine schnelle und zuverlässige Erfassung des Ablenkungszustandes beim Fahrer von hoher Bedeutung sein, wie z.B. zur Entwicklung entsprechender Gegenmaßnahmen oder um die Entstehung solch kritischer Situationen von vornherein besser verhindern zu können. Die EEG-Methode hat das Potential sich als effektive Standardmethode zur Messung mentaler Beanspruchung in der Forschung und Entwicklung zu etablieren und hierüber zur Verbesserung der Sicherheit der Teilnehmer im allgemeinen Straßenverkehr beizutragen.

Summary

Many previous studies have shown that drivers who are under the influence of a secondary task additional to the performance of their driving task can be distracted very quickly–and to a high degree–from the traffic situation which may increase the risk of accident (e.g. McEvoy et al., 2005). Although the EEG has a great potential due to its high temporal resolution as a measurement tool, it has only been insufficiently used to study this research question. In a series of three experiments, that were run in the laboratory and in two real driving experiments, this research gap could be further closed with the work submitted herewith. Over all three experiments, an increase in EEG alpha power was shown which appeared as a consequence of mental workload generated by a demanding listening task. These results were demonstrated in the laboratory both when said tasks were performed alone and when they were performed in parallel to a simulated lane change task. The same alpha band results were shown in real road traffic by using a masked listening task, but not when performing an additionally introduced mental arithmetic task. The third driving experiment replicated the EEG alpha band effects over two different measuring periods by using the same task as in Experiment 1. The consistent correspondence of the results speaks for the high reliability and generalizability of the effects. The observed alpha band increase under mental workload corresponds to the results of previous EEG laboratory experiments studying internalized attention (e.g. Cooper, Croft, Dominey, Burgess, & Gruzelier, 2003, Jensen, Gelfand, Kounios, & Lisman, 2002; Ray & Cole, 1985b) indicating that alpha oscillations are related to the inhibition of task-irrelevant cortical areas. When transferring these results to the context of driving an automobile it could be demonstrated that these results provide a new explanatory approach for the underlying neuronal mechanisms of phenomena such as “endogenous distraction” (Recarte & Nunes, 2003) or “inattentional

blindness" (Strayer & Drews, 2007) that are currently discussed in the literature on traffic psychology extensively.

The additional analysis of ERPs elicited by the stimuli of a tertiary task in Experiment 2 resulted in a workload specific pattern of results that was different between listening and mental arithmetic tasks. For sections with simultaneous mental arithmetic task, an increase in the attention-sensitive N1 component and a reduction in the P3 component were demonstrated. ERPs in sections with a parallel listening task showed a delay in N1 latency of a about 50 ms and a related amplitude reduction which evidently reflected the processing of auditory stimuli concurring at the same sensory channel. These results show, that, in the present case, ERPs were more sensitive to the different workload manipulation by the secondary tasks than the EEG alpha band. However, at this point certain structural flaws in the design of the mathematic task must also be taken into account which could possibly explain the nonexistent effects in the alpha band. Nevertheless, the different results in the ERPs confirm that both tasks addressed different types of workload. Model assumptions predicting interference between incoming signals of the same modality are equally confirmed in this way (Wickens, 1984).

In terms of a multilevel approach (Manzey, 1998) behavioral and ECG data were collected over all three experiments. In Experiment 1, a word detection task integrated into the listening task was compared to a simple tone detection task; whereas, in Experiment 2, the answers to a tertiary task could be evaluated as additional workload measure. A higher error rate and slower reaction times confirm the successful manipulation of mental workload by demanding secondary tasks added while the subject is driving. In the same way, the Lane Change Test in the laboratory experiment shows a higher driver deviance from the ideal change track which also matches already known research results (Mattes, 2003).

The ECG data in the field experiment (Experiment 2) show an increase in heart rate and a decrease in heart rate variability under high workload which is a frequently observed pattern of results reported in scholarly journals and relevant reference literature (e.g. see Ribback, 2003 for summary).On the other hand, the EEG results of Experiment 1 are inconclusive. In fact, there is the same decrease of heart rate variability under higher workload, but also a decrease of heart rate. Moreover, the ECG in Experiment 3 is not sensitive to the workload manipulation due to the secondary task. In general, it has to be considered that these investigations did not include monitoring of respiration. This might have led to confounding variables in the coronary data. Thus, the ECG data are limited in their interpretability. Future investigations should categorically include such respiratory control measurements.

Moreover, a comparison between different baseline conditions (during driving and during rest with eyes closed) and workload conditions revealed multiple interconnections between different cognitive states and changes in alpha power. Based on this insight, a task-specific and individual EEG alpha band adjustment was performed in the work presented. Similar procedures are known from previous laboratory research (e.g. Klimesch, Doppelmayr, Russegger, & Pachinger, 1996). For an application oriented research context, such an approach seems generally recommendable to achieve an optimal signal-to-noise ratio. However, it also became clear that additional research effort is needed to be able to develop theories allowing a better distinction between workload-specific and fatigue-specific alpha.

Finally, it can be concluded that the present work builds a bridge between pure basic-research-oriented laboratory research on the topic of attention on the one side and hands-on traffic psychology on the other side. It has been frequently criticized, that laboratory experiments only implicate an insufficient involvement of the subjects into the task (e.g. Hole, 2007). The stability of the results speaks in favor of the extended usage of the EEG measurement technique in field experiments in real road

traffic. In addition, the frequency band measure generally has the advantage that unlike the event-related P300 measure it doesn't need an artificial extra task. For research and development on driver-vehicle-interaction, the EEG can be of high significance to quickly and reliably assess the drivers' degree of distraction, as for example in order to develop suitable counter-measures or to be able to better avoid such critical situations in the first place. The EEG method has the potential to become an established effective standard method for the assessment of mental workload in research and development and to contribute to the improvement of safety for the participants in general road traffic.

1 Introduction

Driving is a very common task today and its underlying cognitive processes have been widely studied in the field of Psychology. The driving task is a classical example of a man-machine-system[1] that is rarely missing in any classical textbook of Engineering Psychology (e.g. Johannsen, 1993). But driving has also been subject to an intense change over the past decades. Today, the driver is confronted with a variety of potential sources of distraction inside the car. Modern navigation, information and assistance systems enhance the number of available features in the cockpit, e.g. by providing internet access to information such as current traffic conditions or vacant hotel capacities (cf. Krems, Keinath, Baumann & Jahn, 2004). Most current developments even intend to transform the car into a mobile office (see Hole, 2007). It is a more and more accepted point of view that driving time is wasted time that needs to be used productively (Wickens, Lee, Liu, & Becker, 2004). The fact that humans only have limited cognitive resources and that driving is not at all an automatic task can be easily overseen in this context. For example, it is now well known that drivers involved in a secondary task like following a conversation are considerably cognitively demanded which may lead to a significant attentional impairment for their general driving abilities, e.g. to accurately react in case of a sudden, unforeseeable event. In such a situation drivers may be put to greater risks than admitted so far.

The automobile industry has taken great effort in the progression of advanced driver assistance systems (ADAS) to recognize and encounter potentially hazardous situations (see Wiltschko, 2003 for an introduction and brief overview of ADAS).

[1] A man-machine system is defined by the interaction between one or multiple humans and a technical system. Thereby, the term machine is commonly used to describe technical systems of all kind (Johannsen, 1993).

Two questions rise within this context: First of all, how can a state of dangerously high mental workload instantaneously and reliably be assessed? Secondly, which way would be the best to assist an overloaded driver (e.g. adaptive automation, Wickens et al., 2004) or how can potential sources of overload be identified and counter-measures be initiated? To find answers to these important questions, it is inevitable to develop a reliable and fast assessment tool that can detect when the driver is in a state of high mental workload. Moreover, such a tool may also offer the capability for immediate reaction. As pointed out in this thesis, the EEG has important advantages over any other measurement techniques.

From today's point of view, it is out of question that no customer would ever put on an EEG cap before starting to drive a car (e.g. the EEG still requires a significant amount of effort in preparation). But it may be a feasible scenario that autonomous systems are developed that work solely based on vehicle data collected from the CAN bus[2] (e.g. Färber & Färber, 2004). Those systems need careful validation at a pre-developmental stage. To accomplish this, a thorough assessment of the driver via EEG is invaluable. Its application ranges from avoiding system designs that overload the driver to the development of assistance systems that can intervene in highly demanding situations (e.g. so-called workload managers).

Subjective, behavioral and peripheral-physiological methods were used before to address this research question. Unfortunately, as discussed in this work, all of these measures show certain virtues and flaws limiting their application range. On the other hand, neurophysiology may provide crucial advantages: It provides a fast and direct measurement from the driver's brain. Moreover, EEG has been used in

[2] Controller Area Network (CAN) is a broadcast, differential serial bus standard, originally developed in the 1980s by Robert Bosch GmbH, for connecting electronic control units (ECUs) (Controller Area Network, 2007, para. 1).

numerous laboratory experiments to measure working memory and attentional processes. A core finding from this research is that the power in the EEG alpha band increased with increasing cognitive demand and within tasks that require internally directed attention. It has been a challenge for this work to bring a strict laboratory tool like the EEG into an open and noisy environment like car driving and to establish a reliable standard procedure of measurement.

In this way, it was necessary to operationalize a multifaceted task like driving into an experimentally testable scenario that produced reliable and valid results. A set of three experiments was run involving not only EEG data, but behavioral data and data from the electrocardiogram (ECG). Instead of a bottom-up approach in which stimuli are solely a product of naturalistic driving conditions, a top-down strategy has been used for all three experiments. Mental workload has been imposed onto the driver through the presentation of secondary tasks. The presentation of experimental stimuli was consistent from subject to subject, while the variability introduced by other drivers and motorway conditions has been controlled and documented as far as possible.

This work starts with a presentation of theoretical accounts proposed for the context of driving and mental workload. Following this, an overview of already existing empirical studies is provided. Most of theses studies investigated the influence of cell phone use on the driving abilities of the driver. Nevertheless, there has been no published electroencephalographic study before that involved a dual-task paradigm in real-traffic to measure mental workload. Next, the EEG method with a special focus on the EEG alpha frequency band is briefly introduced. Lately, laboratory studies initiated a debate about the functional significance of the alpha frequency band. Some of these findings were used to generate suitable hypotheses for the experiments described in the present work. The results from these experiments are in line with previous EEG laboratory studies and driving studies on cell phone use. It is demonstrated that high mental workload during driving is reflected by an increase in the individual alpha band of the driver. Limits of the EEG method are discussed with respect to testing in artifact-prone environments. The work is concluded by an outlook on limits and potential applications of the findings for research and development in an automotive context.

2 Theoretical Background

The present work focuses on the cognitively loaded driver in real-traffic. Three experiments were conducted in which cognitive workload was imposed onto the driver by secondary tasks. To develop a practical set of testable hypotheses it is fundamental to understand the underlying theoretical concepts of driving and mental workload. A great amount of literature has been published that discussed the different accounts to assess mental workload showing that so far no single technique could be developed to provide a real standard measurement (a so-called "gold standard"). One of the most popular accounts has been the use of secondary task paradigms that led to the development of important theoretical frameworks on drivers' mental workload. Therefore, these paradigms are discussed in a separate section. However, the EEG holds very promising advantages over any other technique and based on its successful previous application (in laboratory and operational environments), it has been chosen as assessment tool for this work.

This section starts with theoretical frameworks known throughout the literature in the context of driving. The most prominent accounts are briefly introduced.

2.1 The Driving Task

2.1.1 Theoretical Models of Driving

Driving is a well studied field and has been described in high detail throughout the literature (e.g. Schulz, 2007). The presented models are meant to provide a basic understanding of the nature of the driving task and any secondary tasks that may be simultaneously performed. Vehicle operation is a classical example of a man-machine system. It includes three main components that closely interact: the driver, the vehicle and the environment (cf. Hoyos, Fastenmeier, & Gstalter, 1995). For the drivers there are two competing goals to pursue: productivity and safety (Wickens et al., 2004). Productivity describes the fact that drivers want to reach their destination in a timely fashion, while safety involves a safe arrival without any accidents.

The driving task has been described by hierarchical three-level models. One of the most prominent models has been presented by Michon (1985). He defined a three-level hierarchy that may underlie cognitive control of driving. It includes (1) the strategic, (2) the tactical, and (3) the sensory-motor control level. On the strategic level the driver decides about the goals of the trip, selects routes and evaluates the costs and risks involved with alternative trips. One level below, the tactical or maneuvering level includes immediate decisions, e.g. choosing a driving maneuver to drive in the speed or driving lane, obstacle avoidance or negotiating a driving situation like making a turn at an intersection. On the lowest level, the operational or control level the driver fulfills mostly automatic, sensory-motor sequences like gear-shifting, steering, or keeping a certain speed. In Figure 1 a slightly earlier and fairly similar hierarchical model proposed by Donges (1982) is displayed in the right part of the figure. Apart from the labeling of the three hierarchical levels (navigational level, driving control level, and sensory-motor level) the three levels are basically identical to the ones proposed by Michon (1985) above.

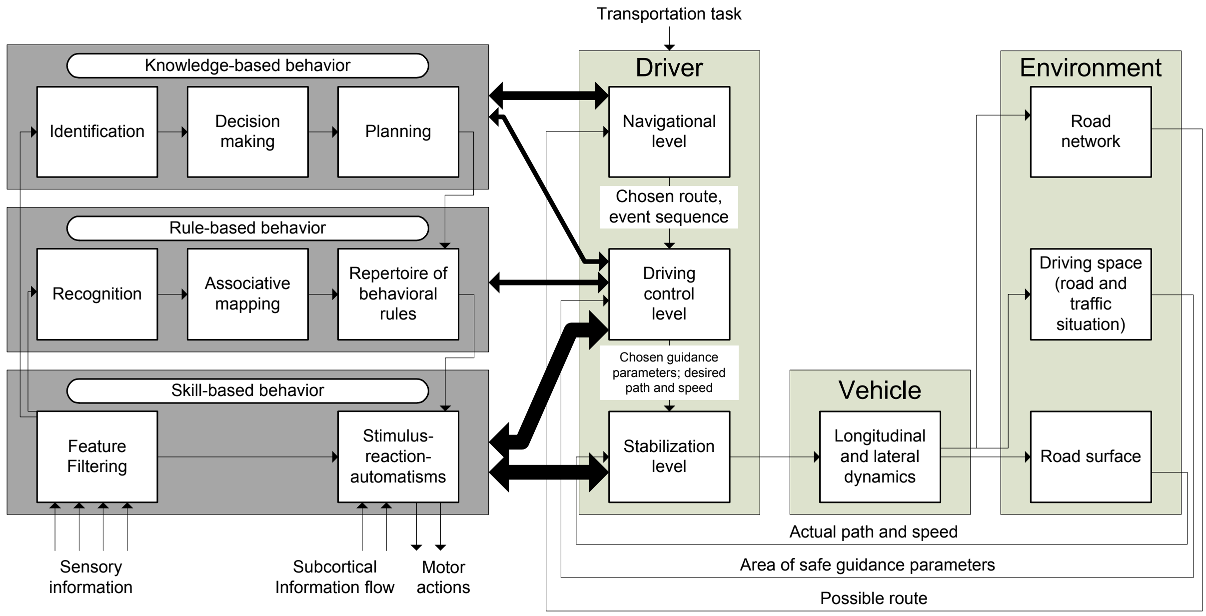

Figure 1. The relationship between Rasmussen's (1983) general model of human behavior (left) and the three level hierarchy model of the driving task according to Donges (1982; figure adapted from Donges, & Naab, 1996, p.227).

Donges and Naab (1996) expanded the model and related the different stages of the driving task to Rasmussen's (1983) general model of human performance. Rasmussen classified human behavior into knowledge-based, rule-based, and skill-based behavior. The higher the level, the higher the complexity of the driving task and the lower the frequency with which the task is usually performed. The knowledge-based behavior is generally invoked in unprepared and untrained situations. It involves complex problem solving and has therefore been assigned to the highest, navigational level. Rule-based behavior involves reactions to situational demands in which rules or productions are automatically activated. This behavior relates to the driving control level. Finally, on the skill-based behavioral level, automated schemata consisting of well-learned procedures are used to unconsciously react to upcoming stimuli. The skill-based behavior can be assigned to the stabilization level. The continuous processing of a constant flow of information has been described by loop control systems at all three levels, e.g. information about the traffic feeds back to the driving control level at which speed and path are adjusted.

The direct assignments on the same level apply for most of the experienced drivers and most of the tasks (Ranney, 1994). However, the authors included arrows which connect levels that are not equal to represent the inter-individual differences in driving experience.

The influence of learning on driving has been studied by Hale and his colleagues (Hale, Stoop & Hommels, 1990). To account for the behavior of novices versus experienced drivers and the performance in familiar versus unfamiliar situations, the authors proposed another relationship between the hierarchical model of Michon (1985) and Rasmussen's (1983) model, which is illustrated in Table 1. Selected tasks are presented to demonstrate behavioral differences based on level of experience and familiarity with the route and vehicle.

Table 1. Examples for the relationship between Michon's (1985) control hierarchy and Rasmussen's (1983) knowledge-rule-skill model.

	Strategic	Tactical/ Maneuvering	Operational/ Control
Knowledge	Navigating in unfamiliar area	Controlling skid	Novice on first lesson
Rule	Choice between familiar routes	Passing other vehicles	Driving unfamiliar vehicles
Skill	Route used for daily commute	Negotiating familiar intersection	Vehicles handling on curves

Note. Highlighted cells represent examples for experienced driver behavior (modified from Hale, Stoop & Hommels, 1990, p. 1383)

The diagonal running from the upper left to the lower right cell in *Table 1* shows how most of the driving tasks cluster for the experienced driver. In this way, knowledge-based behavior is involved at the strategic level, rule-based behavior at the tactical level, and skill-based behavior at the operational level. The other cells show exceptions that point out the differences between familiar and unfamiliar situations and between skilled and novice performance. The example of shifting gears may illustrate the differences: a novice has to consciously remember each step when shifting gears. The inexperienced driver operates on a high, knowledge-based level. On the other hand, an experienced driver is usually not even aware about the process and shifts gears automatically and effortless on a skill-based level. Based on these observations, it can be concluded that the average driver has a high level of driving expertise and that unfamiliarity and lack of practice could seriously confound the results of investigations of drivers' mental workload.

Rasmussen (1983) suggested for his model that although all three levels are operating at the same time, the locus of control, i.e. conscious processing, can only be at one level at a time. Since all three levels are dependent on each other, a change in demands for one level will influence the execution of tasks on the other two (cf. De Waard, 1996). For example, imagine a driver engaged into a difficult conversation on the phone while driving on the highway. Most of the time, the influence onto the skill-based level can be easily observed, since the driver drives slower and reacts slower to situational demands. The differentiation of different levels of the driving task has been an important theoretical advance, because it allows a fine grained analysis of the influence of mental workload on each level. For the empirical work reported in this thesis, degradations in the drivers' performance due to the imposed mental workload should most likely be observable on the operational and tactical level. However there have been some arguments from the research on automaticity going even further. In the following section, evidence will be presented indicating that driving performance is hardly ever completely autarkic even at the most elementary level.

2.1.2 The Role of Automaticity in Driving

As early as 1938, Gibson and Crooks identified automaticity to play an important role in driving that needed further investigation in driving research. Today, automaticity is known to be a central construct in cognitive psychology that has mostly been studied in the context of elementary processes such as detection and memory search (Schneider and Shiffrin, 1977). Automaticity develops following extended consistent practice and it refers to fast, effortless processing (Schneider and Shiffrin, 1977; Shiffrin and Schneider, 1977). Moreover, automated processes are self-paced and with only little recollection of specific elements of the task. They can be performed alongside other tasks and rely on open-loop control, i.e. no feedback from preceding actions is needed. Automated processes have been contrasted to controlled processes which have been characterized as slow, serial and effortful. The results from this research have often been applied to the driving task. As mentioned in the previous section above, under normal conditions the operation on the tactical and the stabilization level shouldn't require attention at all times. Most of the tasks at these levels are known to be bottom-up tasks that are highly automated and in which sensory information is processed without awareness. Based on these observations, one might conclude that an automatic task like driving or more specifically the lower levels of the task should hardly be affected by any additional cognitive task.

Nevertheless, Groeger (2000) argued against the expansion of any rigorous account of automaticity to a complex task like driving. In his book, he reviewed research studies that examined the cognitive costs of gear shifting (Duncan, Williams & Brown, 1991; Groeger & Clegg, 1997). The results of these studies showed that basic principles of automated tasks obviously did not apply for this low level task: (1) performance differences in novices as well as in experts indicated that experience level was irrelevant for the automatic performance of gear shifting and (2) a considerable amount of variation in timing costs for shifting gears could be demonstrated that should not be observed in an automated task. Groeger concludes:

“I believe that gear changing offers the most likely opportunity for automaticity to be demonstrated in the driving task. … the evidence is that gear changing is not automatic implies to me that very little, if any, of the driving task is [automatic] – and beyond that, that we should seriously doubt the utility of strength based accounts of automaticity.” (Groeger, 2000, p. 69). In his opinion, it is more reasonable to see driving as a broad range of activities that comprises a multiple set of capacities (Wickens, 1984; see multiple resource model below) along with well practiced routines building a single control architecture. Although Groegers’ conclusions are not shared by everyone in the research community (cf. Underwood, 2002), he emphasized the fact that at each level of the driving task resources can be bound.

2.1.3 Environmental Factors

The environment represents a complex part in the man-machine system of driving. Bernotat and Käppler (1985) differentiated between the natural, the configured, and the social environment. The natural environment includes all elements that are given through nature, i.e. terrain, daytime, wind, and weather. The configured environment characterizes every element constructed by humans to assist in the transportation task, e.g. the road network or the road design. Finally, the social environment encompasses all the other road users and aspects like traffic density and traffic flow. The influence of the environment onto the driving system has been subject to intensive research (Hering, 1999; Verwey, 2000; De Waard, 2002; Schmidt, 2006). Based on previous work by Benda, Hyos and Schaible-Rapp (1983), Fastenmeier (1995) presented a systematic taxonomy of driving situations which allowed a prediction of driving complexity. Through combination of the different elementary categories a reconstruction of most of the existing street sections was possible. Fastenmeier employed a questionnaire (Fragebogen zur Arbeitsanalyse

(FAA), Frieling & Hoyos, 1978) that has initially been developed for general work analysis and used it to assess the complexity of the different situations. As a result, Fastenmeier presented indices of complexity that could be used as a predictor for mental workload in driving. As subsequent empirical works have shown, Fastenmeier's taxonomy provided a fairly good predictability in this respect (e.g. Harms, 1991; Verwey, 2000; Brookhuis & De Waard, 2001; Schmidt, 2006).

Rauch and colleagues (Rauch, Totzke, & Krüger, 2004) critically reviewed various accounts (e.g. Fastenmeier, 1995; Hering, 1999; Schumacher, 2001; Verwey, 2000) that tried to estimate mental workload based on driving situation, driving tasks and accident report analyses. As they explained, none of the existing accounts has been comprehensive and the most promising idea would be to assign a basic level of task demand based on static influences (e.g. road condition, road type etc.) modified by dynamic aspects, e.g. sudden braking of the preceding car. In general, the authors criticized the lack of consensus in the methodological approaches that were used in the previous studies, i.e. there were real traffic experiments or simulator studies that used various measures such as driving performance, psychophysiological, and subjective parameters as well as secondary task performance. Most of the studies used different reference standards. As Rauch and colleagues pointed out, with each methodological account a different aspect of mental workload is measured and therefore no absolute rating of task demand could be achieved. In their review, they came to the conclusion that four essential factors of demands during driving can be identified: (1) road features, (2) driving maneuvers, (3) traffic influences, and (4) environmental conditions or visibility conditions. By simply adding up the classification results of the reviewed studies ($N = 13$) they attained a coarse overview of the most prominent factors. The most demanding situations were: driving in curves, driving on a lower order street or residential area, intersections where the driver has to yield right of way, turn left and right, highway entrance and exit, driving on a highway with noise protection walls and passing a tunnel. Less demanding situations were: driving on rural roads and highways, driving on a higher order street inside a town, approaching intersections with right of way and driving on the freeway. Highly demanding driving maneuvers were: overtaking, turning, unexpected braking and lane changes. An increased demand based on traffic stemmed from interactions with other traffic participants and high traffic density. Finally, environmental factors such as bad visibility (e.g. fog or heavy rain) or bad road conditions (e.g. wet or icy roads due to rain or snow) additionally increased driving demands.

The research reported above demonstrated the impact of environmental factors on driver's mental workload. In contrast, the empirical work of this thesis focused on drivers' cognitive workload imposed by secondary tasks. The assessment in a real-world setting has been essential and inevitable for reasons discussed in detail below (see section 2.4.2.2). The profound influence of the driving environment onto drivers' mental workload needed to be carefully considered in the planning of the two real-traffic driving experiments of this thesis. Nevertheless, please note that even with the best experimental design and instructions, a field experiment will never reach the same degree of control as a laboratory experiment. Research results describing the driver as the central part of the man-machine system driver-vehicle-environment are summarized in the subsequent section.

2.1.4 Neural Correlates of Driving

Most research describing the individual performing the driving task has been published in the field of traffic psychology which aims at the assessment of a driver's ability to drive. The high variety of brain structures and cognitive processing capacities that are involved in the highly complex driving task become evident when considering cases of brain injury. It is well known, that recovery from traumatic brain injury is always incomplete and that the resumption of everyday activities like driving may be problematic due to deficits in remembering, organizing, learning, and planning (Groeger, 2002). Until today, the relationship between standard psychometric tests and driving performance is not fully understood and usually an extensive assessment of these patients has to be performed. Groeger listed three criteria that have to be ensured before being able to confirm an individual's fitness to drive: (1) a very low probability of sudden and unpredictable lapses of control over behavior, (2) sufficient perceptual, cognitive, and motor abilities, and (3) sufficient social judgment and social responsibility. Especially for older people the existing neuropsychological tests (e.g. see McKenna, Jefferies, Dobson, & Frude, 2004 for a battery of cognitive tests) have been considered insufficient and some authors (e.g. Groeger, 2000) pointed out the necessity to always include tests that are as similar to real driving conditions as practicable.

Further insights have been expected from brain-imaging studies investigating the neural correlates of driving. Walter et al. (2001) ran a fMRI study to investigate the activation of brain regions during simulated driving. They found an increased activation in the left sensorimotor cortex and in cerebellar regions that would most likely be associated with driving, i.e. controlling the car via joystick with the right hand. Moreover, increased activity was found in the right occipital and parietal regions. These two areas seemed most likely linked to perceptual processes during driving. Against the authors' predictions the area MT/MST showed a bilateral deactivation during active driving compared to passive driving. According to Walter and his colleagues, this might have come from an increased optical flow when watching the drive or it might relate to an attention modulation process, i.e. a suppression of MT/MST activity due to a high focus on controlling the car. Passive driving also led to a higher activation of the frontal eye fields (FEF) and an increased activation in the right dorsolateral prefrontal cortex (DLPFC). The former seemed to relate to increased eye movements, the latter was concordant to a more active navigation throughout the course as reported by the subjects. Based on these results, the authors concluded that a general limited capacity model of driving would be inadequate to capture the complex neurobiological processes of driving. They demonstrated that while motion sensitive areas were less active, the coordinated activity of occipito-parietal and motor brain areas was required for simulated driving.

A consecutive fMRI study (Calhoun, Pekar, McGinty, Adali, Watson, & Godfrey, 2002) decomposed the neural activity during simulated driving into independent components. They confirmed the results of Walter and his colleagues by demonstrating activation of cerebellar and occipital areas related to visuomotor integration. Moreover, they found that decreases in activity of the anterior cingulate cortex and frontoparietal regions were associated with higher speed in simulated driving. The authors assumed that these regions would be generally associated with a change in vigilance as well as error monitoring and inhibition processes.

In a more recent PET study (Horikawa et al. 2005), the authors took a closer look at correlations between regional cerebral blood flow (rCBF) and the time required to complete a course as well as the number of crashes. Besides activation related to visual information processing in the parietal-occipital regions and motor control processes in the activated cerebellum, Horikawa and colleagues showed that impaired driving was correlated with increased rCBF in the thalamus, midbrain, and posterior cingulate gyrus. The latter has previously been shown to be related to error-related brain activation during driving tasks (Menon, Adleman, White, Glover, & Reiss, 2001). From animal studies and an earlier PET study, it is known that thalamus and midbrain are involved in processes of attention and high vigilance (Kinomura, Larsson, Gulyas, & Roland, 1996) as well as "interval timing", i.e. estimating the time needed to complete a task (Hinton & Meck, 1997). Therefore, the authors concluded that the activation in these brain regions would be closely associated with maintenance of driving performance. In their outlook, the authors were optimistic that functional neuroimaging may provide useful information for assessment of traffic accident risk.

In line with previous results, brain imaging studies demonstrated the complexity of the driving task involving the multifaceted interconnections of multiple brain regions that underlie a variety of cognitive processes of perception, attention, orientation, decision-making, and motor skills. However, solely based on the results of these studies a direct mapping of the intensive interplay between single operations and the underlying brain regions has not yet been possible. More research is needed to allow the useful application of these results, e.g. to assess driving skills or to measure the impact of distraction on driving ability. However, any simulator study in a laboratory environment has the significant drawback that it mirrors only a very limited part of the real driving situation. The driving simulation resembles "more a training run of a race driver than an everyday driving situation" (Walter et al., 2001, p. 1766).

Although research studies have demonstrated the high complexity of the driving task, it has been a general conception in the population that driving is a boring task. For that reason, drivers often look for some ways to activate themselves. The first thing at hand could be to talk to passengers in the car. If there are no passengers, most of today's cars have a radio or some people use cell phones during driving. As a study by Nordbakke and Sagberg (2007) reported, cell phone use might be regarded as a possible countermeasure against sleepiness among today's drivers. However, under high mental overload additional activities during driving could quickly endanger road safety. Keeping the car on the road may be a fairly automated skill, but in a critical situation a conscious interference may become necessary, e.g. to avoid a sudden obstacle on the road. If the driver has no free capacities available at this moment to react, an accident may result. The following review of workload models shall help to further understand how the driver manages limited cognitive resources to simultaneously accomplish driving and additional tasks in parallel.

2.2 Theoretical Models on Mental Workload

Mental workload is a miscellaneous term that is commonly referred to in automotive research. Despite its extensive use, there is neither a common understanding of the term “workload” nor a shared methodology for measuring it, especially with driving. Historically, the concept of “mental workload” was first introduced in the 1940s in the context of optimizing human-machine systems (Bornemann, 1942 cited after Manzey, 1998). In more recent publications, the overall workload of a person is regarded as a composition of physical, emotional, and mental workload components (e.g. Gaillard, 1993; Ribback, 2003). For the driving situation, the mental and emotional components are more relevant than the physical component. To properly distinguish mental and emotional workload it has been proposed that mental workload solely refers to task-specific demands that are imposed onto the human information processing system. In contrast to that, emotional workload refers to demands imposed during action execution, i.e. time pressure, noise, heat, danger and social conflicts that are related to feeling of anxiety and helplessness. This classification is far from being sufficient from a theoretical point of view, but it offers a first pragmatic description of the term “mental workload” (cf. Manzey, 1998).

2.2.1 Mental Workload as Defined by ISO 10075

To accomplish a standard terminology for the concept of mental workload the international norm ISO 10075 has been created (Deutsches Institut für Normung e.V., 2000; International Organization for Standardization, 1991, 1996, 2004). The norm was developed to provide a theoretical fundament as well as hands-on guidelines for applied work in health and work psychology. The norm contains three parts. Part 1 is concerned with the terminology and concepts of mental workload, Part 2 encompasses guidelines for work design and Part 3 focuses on the requirements for the development and application of workload measurement methods. The definition of mental workload described in ISO 10075-1 is identical to the previous definition in the German norm DIN 33405 that was published in 1987 (Deutsches Institut für Normung e.V., 1987). Mental workload has been defined according to Rohmert's (1984) distinction between mental stress and mental strain that originally stems from the field of mechanics. The authors of ISO 10075-1 defined the term "mental" to refer to cognitive, informational, and emotional processes in the human being. Mental stress has been defined by ISO 10075-1 as: "The total of all assessable influences impinging upon a human being from external sources and affecting it mentally". (International Organization for Standardization, 1991, p. 1). Therefore, mental stress describes all external and psychological factors that influence an individual. This includes aspects of task demands, work equipment, physical work environment, and social work environment. Whenever an individual encounters these influences mental strain is the consequence. Mental strain is defined as: "The immediate effect of mental stress within the individual (not the long-term effect) depending on his/her individual habitual and actual preconditions, including individual coping styles." (International Organization for Standardization, 1991, p. 1). The person's reaction to these external factors is completely dependent on personality, qualification, and current capabilities. The direct consequences of mental strain may be mental fatigue or fatigue-like states on the one hand or facilitating effects (e.g. "warming-up effects" or activation) on the other hand. Training effects have been mentioned as possible indirect consequences of mental strain. The term "stress" in this norm is used as a completely neutral term. Due to the negative bias of the word "stress", Nachreiner and Schultetus (2002) recommended to replace it with the term "mental workload". Following this, the terms "mental stress" or "mental workload" would focus completely on factors outside the individual. This view may be inadequate if other existing models of mental workload throughout the literature are considered (see below). Moreover, the two component model described in the ISO 10075-1 norm has been criticized to oversimplify the concept of mental workload. But after all, the account's simplicity holds some advantages, e.g. it has been successfully incorporated in the German national collective bargaining agreements (cf. Bamberg, 2002).

2.2.2 Further Models of Mental Workload

In the literature on mental workload, a number of theoretical models have been presented that highlight different aspects of the concept. Some of the most relevant concepts have been selected and are presented in the following (see Manzey, 1998 or Ribback, 2003 for more comprehensive reviews).

Some authors (e.g. Gelau, 2004) recently claimed that the term driver mental workload could be used synonymous with driver distraction. In lack of a suitable definition, Gelau adapted the stress-strain concept of Rohmert (1984, see text above), as it has been used for mental workload, to define the consequences of in-car information- and communication technology onto the driver. On the other hand, he clearly distinguished driver distraction from inattention. This was in agreement with Sussmann, Bishop, Madnick and Walter (1985) who stated: "…inattention may be the result from introspective behavior by the driver or a distraction by the driver. The operational result is that the driver makes a delayed response, an inappropriate response, or no response at all." (p. 41).

In the present work the consequences of demanding secondary tasks on the driver's cognitive state have been assessed. In contrast to Gelau's interpretation other concepts showed that the term mental workload is more than just simple distraction. It refers to a mental state that reflects the consequence of external task demands. For this thesis, a conception of mental workload has been used that encompasses a variety of aspects as presented in the following frameworks.

2.2.2.1 Fundamental concepts of mental workload.

A central concept for mental workload is the notion of a limited amount of cognitive resources (Kahneman, 1973). Within this concept the amount of mental workload is equated to the amount of capacity resources needed to perform a specific task. In this way, the level of mental workload is inferred from the currently available spare capacity of the system. It has been assumed that performance would decrease if task demands exceeded those resources, but mental effort could be recruited to meet the increased requirements. Thus Kahneman assumed that the total amount of capacity is not constant. With respect to his idea it has been frequently criticized that the definitions of the terms "resources" and "mental effort" are very vague: "Resources could be envisaged in terms of neural processing capacity (neurons can only do so much at a time, because they are limited by their maximum firing rate), but this probably only replaces one vague term with another!" (Hole, 2007, p.64). Nevertheless, the idea of "limited resources" has been a fundamental concept for all subsequent models.

O'Donnell and Eggemeier (1986) presented their model on mental workload in which they defined workload to be the part of limited cognitive resources needed to fulfill a certain task. According to the authors, workload maps directly to the amount of task difficulty. Workload during driving is regarded to be task dependent, i.e. it is determined by driving situation, vehicle instrumentation, degree of automation and feedback as well as characteristics of the driver (e.g. trait variables like driving experience, age, strategies or state variables such as degree of monotony or fatigue and the influence of drugs or alcohol).

A different view that related to the aforementioned distinction between automated and controlled processing (see chapter 2.1.2) was published by Shallice and colleagues (Norman & Shallice, 1986; Shallice & Burgess, 1993). In their model of executive control they distinguished between a lower level contention scheduling system and a higher level attentional system. Various patterns of actions that add to the completion of a task are represented in schemas. Multiple schemas compete for their execution and most of the time the contention system routinely decides on an unconscious level which schema gets activated. This decision is influenced by perceptual input as well as the person's current motivational level. However, the voluntarily controlled supervisory attention system can intervene and prime adequate schemas or inhibit unsuitable schemas if this is required by the current situation. Predictions in case of overload are similar to the previous model described: all task demands either from the primary driving task or the secondary workload task are directly addressed to the supervisory attentional system that has to cope by activating appropriate schemas. If task demands exceed the available limited resources the conscious upper level system will fail and performance will deteriorate.

So far all presented models embrace the two concepts of limited cognitive resources and unconscious or conscious processing levels. In general, task demands seem to affect the conscious upper processing level which is in charge to manage limited resources first. With respect to resources, Wickens (1984) contributed a crucial aspect to existing models. His model incorporates qualitatively different types of mental resources and has been widely recognized (see *Figure 2*).

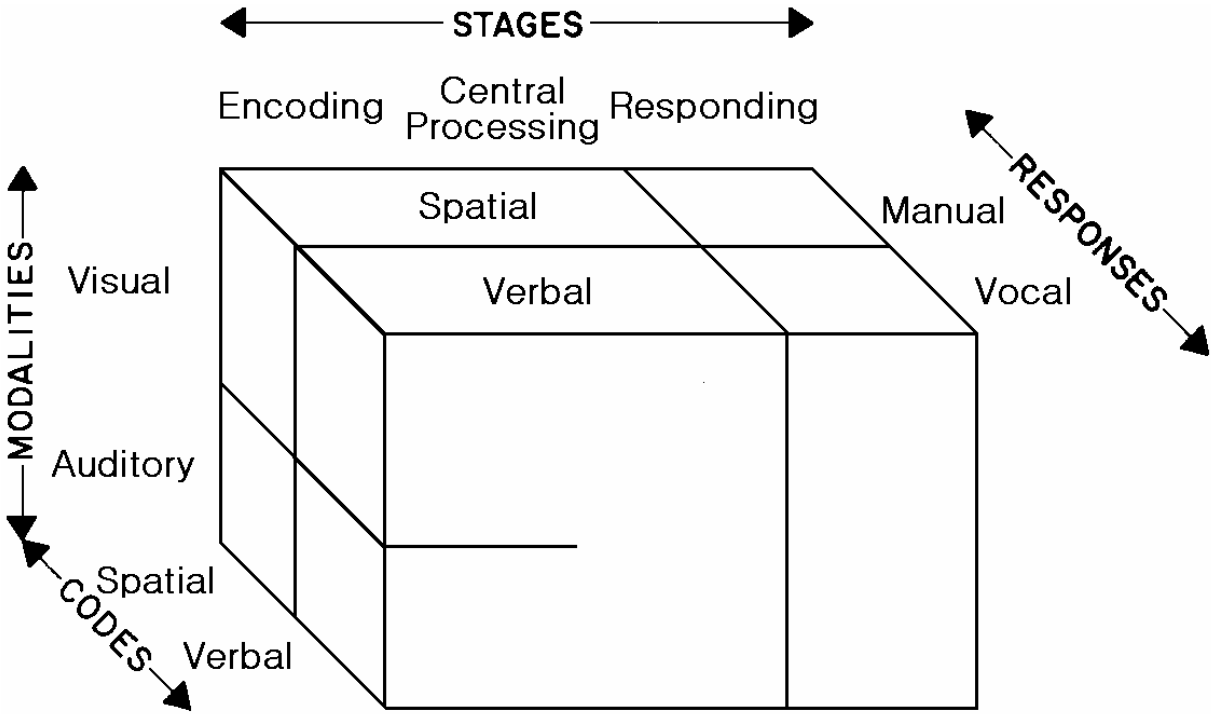

Figure 2. Structure of processing resources according to Wickens' multiple resource model. Operations on either side of a solid line use different resources (adapted from Wickens, 1984, p. 81)

The multiple resource model includes three different dimensions that are elementary to information processing: (1) the mode of information input (auditory, visual, or tactile), (2) the information code (spatially or verbally), and (3) the type of response (manual or vocal). According to Wickens, conflicts arise when the same type of dimension is simultaneously addressed, e.g. a driver who talks to a passenger will have problems receiving the message of an incoming radio traffic report at the same time. In this situation, two pieces of information interfere with each other in the same auditory input modality. In addition, the model describes three processing stages: encoding, central processing, and responding. In terms of the multiple resource model driving can be described as addressing visual input, a spatial code, and manual output such as steering. In contrast, a secondary task like talking to a passenger would involve auditory input, a verbal code, and vocal output. If these two tasks are performed in parallel usually no conflicts would be expected. Thus, for an experienced driver being engaged into a conversation while driving should usually be no problem. This observation has also been used as an argument against Kahneman's idea of a general unspecific resource pool. However, there is another potential source of interference between the two tasks. Performance impairment would also be expected if primary and secondary task both access resources that lie on the same central information processing level (cf. Nunes & Recarte, 2002). As described in more detail in section 2.4.2.2 below, the authors found that there is no difference between talking on a hands-free phone and talking to a passenger, but that it is the conversation content and its complexity that are the real potential distractors.

2.2.2.2 Feed-back control models of mental workload.

Humans employ different strategies to adapt to changing task demands. For example, especially in dual task paradigms there is always the possibility that drivers set their own priorities for the two tasks independently. That means, although having been instructed that driving should be of first priority, it is impossible to ensure that drivers keep these priorities as instructed throughout the experiment. This is why observational data can't provide a direct representation of the current amount of stress. Only the result of task demand, perceived workload and compensational mechanisms can be measured. Two prominent accounts have been known throughout the literature that went beyond a simple cause and action model and that described the stress-strain concept within a feed-back control system.

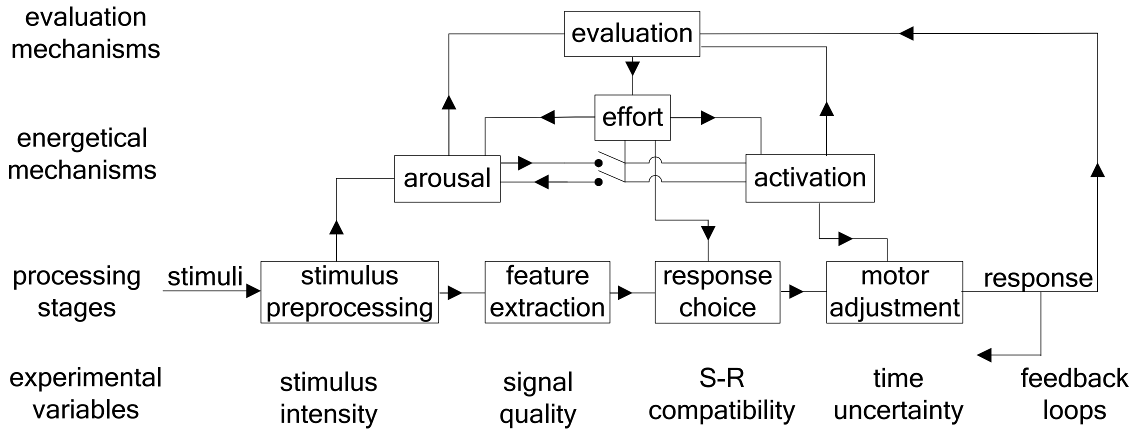

Figure 3. A cognitive-energetical stage model of human information processing and stress (adapted from Sanders, 1983, p. 79)

Sanders (1983, see *Figure 3*) distinguished four levels of human information processing: (1) stimulus preprocessing, (2) feature extraction, (3) response choice, and (4) motor adjustment. The stimulus preprocessing level is automatic and passive and does not require any energetic resources. The remaining three levels all involve active and controlled processes that each requires a unique pool of energetic resources which is in accordance with Wicken's (1984) multiple resource model. For Sanders these energetic resources are equivalent to different types of attention. An operator employs selective attention to distinguish relevant and irrelevant features. The response choice involves conscious decision making processes. Finally, the motor adjustment is optimized through preparatory processes and adapted timing. In the model three energetic systems are described: an arousal system, an activation system, and an effort system. The arousal system represents the phasic reaction to an input, while the activation system reflects a tonic readiness to react. The effort system functions as a management instance that coordinates the activity between arousal and activation. The effort system's main responsibility is to influence conscious decision making by which it contributes to the efficacy of information processing. Furthermore, it can interrupt the direct link between activation and arousal. The model also includes a control instance that Sanders called the "evaluation" instance. It receives a constant feedback about the physiological state of the system and employs the effort system if an imbalance is perceived. It also controls the current level of performance and activates the effort system in case of a discrepancy between current and wanted performance. Sanders distinguished automated and controlled processing within the model. Automated processes work through the direct connection between arousal and activation. The response choice is automated and as long as these processes are successful no activation of the effort system is needed. The effort system is only employed if discrepancies occur that require a conscious response choice. It represents the core compensation mechanism in this model. Effort supervises arousal and activation and it aims at the maintenance of a certain performance level. If this maintenance of performance is not achieved over a longer time period the feeling of mental workload is the result. Please note that employment of effort is not generally causing strain. This is only the case if the effort system continuously fails to establish a balance between activation and effort for a longer time. The reasons for such a failure may lie in an overly high or low activation or arousal. It can also be that even with the greatest effort no decision could be made between ambiguous inputs.

Sanders' model has been recognized as being of high heuristic value (Manzey, 1998). It addresses various energetic resources and describes them in high detail. Moreover, the hierarchical structure of the model and the differentiation of specific energetic mechanisms show that the effects of mental workload can be described on different levels. The effort level encompasses the rather unspecific energetic resources. Here, central processes involved in complex task performance and changes in activation based on motivational processes that are also related to the subjective feeling of effort are addressed. A second level of specific energetic resources (arousal, activation) allows addressing workload related to reaction-specific motor processes and processes of information intake. Nevertheless, the model fails to describe the different types of compensation in more detail. A second compensatory model seems to be complementary to Sanders in this respect.

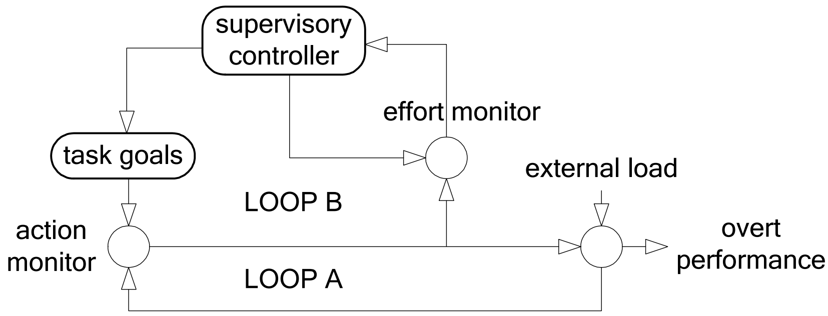

Figure 4. Compensatory control model of performance regulation . Loop A represents routine regulatory activity, and loop B effort-based control (adapted from Hockey, 1997, p. 79)

As previous models, Hockey (1997, see *Figure 4*) assumed a limited pool of multiple cognitive resources and simultaneously performed mental operations that involve similar resource demands. He also suggested the existence of routines that are generally automatically performed at low cost and that involve well-learned skills. At this automatic, so-called "action monitor" level the individual compares the current performance with the desired outcome and the performance is automatically adjusted until the discrepancy between desired and actual outcome is within tolerable limits. As Sanders' model, this model has been expanded to include a motivational energy-regulation mechanism that can compensate for psychological and physiological costs. This compensatory mechanism is represented by the "effort monitor". Its employment is usually perceived as unpleasant and may sometimes even lead to anxiety. Besides providing activation resources, mental effort also helps to maintain the sufficient motivation to achieve behavioral goals and to defend them against competing goals. Mental effort is consciously managed by the individual who constantly has to take goals, task demands, and energetic resources into account. Effort is not automatically employed. In fact, changes in task demands have to be consciously perceived which then leads to a switch of control to the higher feedback control system, namely the "supervisory controller". Basically two different types of compensation can be distinguished which have different consequences for the performance and the costs for resources: During active compensation the maximum effort is increased to manage the increased demands. Thus performance can be maintained, but costs for resources are increased. On the other hand, a passive compensation can be employed during which the expectations for the outcome are decreased. For example, this may be achieved by decreasing the expected accuracy or speed with which a task is completed or by neglecting less important tasks. In this way, the amount of effort is kept constant and no additional costs are involved.

The models of Hockey and Sanders share the same basic assumptions. Both assume the existence of comparison processes between actual performance and the desired outcome (evaluation process and action monitor). If a discrepancy is perceived an effort system is engaged to compensate for higher task demands. Effort involves increased costs of resources and an increased experience of strain. Hockey's and Sander's theories complement each other: Sanders included a detailed description of the different processing steps, while Hockey discussed compensation strategies.

In sum, the presented models give a coarse overview about prominent theoretical concepts on mental workload during driving. To some extend these models provide an understanding of how a driver is able to manage all incoming information with only a limited pool of cognitive resources available. Predictions are possible and allow speculations about potential conflicts in information processing channels e.g. as it has been demonstrated in the design of navigation systems (Verwey & Janssen, 1988, cited after Gstalter & Fastenmeier, 1995). However, theses models have also been criticized for assuming a general pool of cognitive resources without further specification. It has to be admitted that until today, these theories provide only a very vague framework for cognitive processes. Neuroscientific research has produced some intriguing results with respect to attentional processes and task demands that should be used to develop a more comprehensive understanding of driving under high workload.

2.3 The Assessment of Mental Workload

A myriad of techniques for the measurement of mental workload have been developed, however up-to-date there has been no “gold standard” established. In the present work several assessment techniques have been used that all have specific advantages and disadvantages.

2.3.1 Criteria for the Evaluation of Assessment Techniques

For a qualified evaluation of the empirical results it is important to have adequate evaluation criteria at hand. O’Donnell and Eggemeier (1986) defined a set of five criteria for the evaluation of workload assessment indices which has been frequently cited throughout the literature. A few years later this set was expanded by Wickens (1992) to an overall set of seven criteria which need careful consideration in every serious investigation in this field of research.

Sensitivity:

A measure is sensitive if it allows reflecting changes in workload during the performance of tasks with different levels of complexity. Sensitivity is usually investigated by the variation of task demands and task complexity on multiple levels and the observation of changes in the selected variable. For this criterion to be valid it is an absolute prerequisite that the subject is motivated to perform well in the task.

Diagnosticity:

This criterion describes whether an index reacts selectively to a specific type of demand and therefore allows conclusions about the cause for a certain workload variation. With regard to prominent theoretical accounts on mental workload, Manzey (1998) further contrasted diagnosticity into a wider and narrower sense. Diagnosticity in a wider sense refers to the differentiation of physical, emotional, and mental workload. An index may be regarded to be diagnostic in a more specific sense if it addresses the type of resource at which the effects of task demands are displayed. This could either be a global level like the multifunctional effort level described by Sanders (1983) or it could be a more specific level like a certain stage or input modality level as defined by multiple resource theory (Wickens, 1984). Please note, that some authors (De Waard, 1996; Wickens & Hollands, 2000) consider the latter description as the only definition of diagnosticity, i.e. a technique is high on diagnosticity if it reflects specific resource demands and it is low on diagnosticity if it reflects general demands. For the present work the definition by Manzey seemed most useful. Generally, a high diagnosticity may be particularly useful when it comes to the design of countermeasures against high workload demands.

Selectivity:

This criterion has not been listed by all authors. According to De Waard (1996) it defines the validity of a measure for workload assessment, i.e. a selective measure for mental workload is only sensitive to differences in resource demand and it should not be affected by factors such as emotional stress or physical load. For Manzey (1998) this criterion is simply considered to be a part of diagnosticity (see paragraph above).

Obtrusiveness:

A measure should not disrupt, contaminate or interfere with an ongoing primary task performance. This criterion is especially important when operators are performing their tasks in a real world environment in which a failure in task performance may be safety critical (e.g. flying an airplane). While physiological and self-reports taken after completion show the lowest level of intrusion, secondary tasks seem to be most likely prone to interference.

Bandwidth and reliability:

A measure of mental workload should provide a reliable estimate for both within and across tests. Within this context transferability (e.g. Wierwille & Eggemeier, 1993) plays an important role, i.e. a technique developed in the laboratory may not always be applicable in the field as well. Moreover, in some contexts it may be essential that a measure offers a reliable estimate of workload rapidly enough so that appropriate actions can be taken (workload-adaptive systems, e.g. Michon, 1993; Piechulla, Mayser, Gehrke, & König, 2003; Kohlmorgen et al., 2007).

Implementation requirements:

This criterion refers to very practical constraints like the amount of equipment that is needed for the assessment and the amount of practice that an operator needs to perform a task to reach a reasonable and stable performance. For example, in secondary task designs this may be an important factor if extensive training time is required before measurements can be taken.

Operator acceptance:

The correctness and accuracy of a measure may be affected by the operator's opinion about the technique. For example, if an operator does not approve the use of a self-report the accuracy and correctness of the measure can be largely affected. Higher levels of acceptance can be obtained by using measures showing high face validity. Moreover, the less a measure is intrusive or artificial, the better its approval will be. O'Donnell & Eggemeier (1986) suggested including a sufficient amount of explanation about the measurement's use and usefulness at the beginning of an assessment. Acceptance in a secondary task design can also be improved by the use of embedded secondary tasks that are already part of the primary task, i.e. measuring the number of rear-mirror glances during car driving (Wickens & Hollands, 2000).

Generally, sensitivity and diagnosticity are considered being the most important criteria, while the other criteria may be considered additional selection criteria (De Waard, 1996). As several authors (De Waard, 1996; Wickens & Hollands, 2000) pointed out, the different criteria are not independent from each other and some criteria may trade off with one another. Thus there is merely a technique that can satisfy all criteria at the same time. The selection of a certain workload index may mostly depend on the measurement's objective.

In the second part of this chapter, different workload assessment techniques shall be reviewed with regard to the above described criteria. Techniques have been generally classified into four coarse categories that are primary-task measures, secondary-task measures of spare capacity, physiological measures, and subjective rating techniques (Wickens & Hollands, 2000). Tsang and Wilson (1997) added another category, the analytic methods which refer to modeling efforts for data prediction and evaluation. Please refer to their overview article for detailed information on this specific type of technique.

2.3.2 Primary Task Measures

The first measure at hand is the operator's performance on the task of interest. With an increasing level of workload additional processing resources are utilized and it is assumed that this will affect the quality of operator performance (usually degradation). Measuring the changes on performance should provide an index of the workload of the task. The outcome should be a highly valid measure since it directly maps the result of the operator's efforts (O'Donnell & Eggemeier, 1986). However four major drawbacks of this account have been identified. First, in case of low task demands primary task performance is an insufficiently sensitive measure since an increased task demand can easily be compensated by higher effort investment leading to no observable performance degradation. On the other hand, in cases of overload performance changes may already have reached a certain limit where no additional changes can be observed. Therefore, primary task measures show their highest sensitivity at an intermediate level of workload. Secondly, by using this technique a comparison between two different primary tasks is hardly possible. A primary task measure is unique to its task, i.e. two tasks may differ in how they are measured and how the outcomes are interpreted. There is a clear need for measures that are not task specific. Thirdly, some tasks may simply not allow the recording of good, continuous measures for primary task performance. For example, a complex cognitive task with only a single outcome would not reflect the amount of cognitive processes that were needed to obtain the result. Finally, when comparing two different primary task performances it is often the case that differences are caused by reasons unrelated to workload. For example an operator may show a different performance simply due to a different task performance strategy.

For the reasons mentioned above, primary task measures have rarely been used as the only data source for mental workload assessment. Other measures may allow a more direct way to assess either the effort invested or the level of residual capacity that is still available.

2.3.3 Secondary Task Measures

When using a secondary-task assessment technique, two different standard paradigms have been recommended to measure dual-task performance (cf. O'Donnell & Eggemeier, 1986), namely the "subsidiary task paradigm" and the "loading task paradigm". From the following description, it becomes evident that both paradigms require a profound knowledge of the underlying structure of the examined tasks and that the correct selection of the paradigm is essential to being able to answer any predefined hypotheses.

The subsidiary task paradigm is based on the assumption that it can be used as a measure of spare capacity that is not utilized by the primary task. Provided that the secondary task is sufficiently demanding, an inversely proportional relationship between secondary task performance and primary task resource demands has been assumed. Early accounts postulated a totally undifferentiated capacity that is used by all tasks (De Waard, 1996). With regard to multiple-resource theory (Wickens, 1984) it can be assumed that the sensitivity of the secondary task measure is high if the overlap in resource is high, too. That means that ideally secondary and primary task should draw upon the same type of resource. When applying this paradigm, the operator is asked to perform as well as possible on the primary task and spend any residual capacity available on the performance of the secondary task. Within this paradigm, the primary task is the task of interest and its priority is emphasized. The idea of the concept is illustrated in *Figure 5a*.

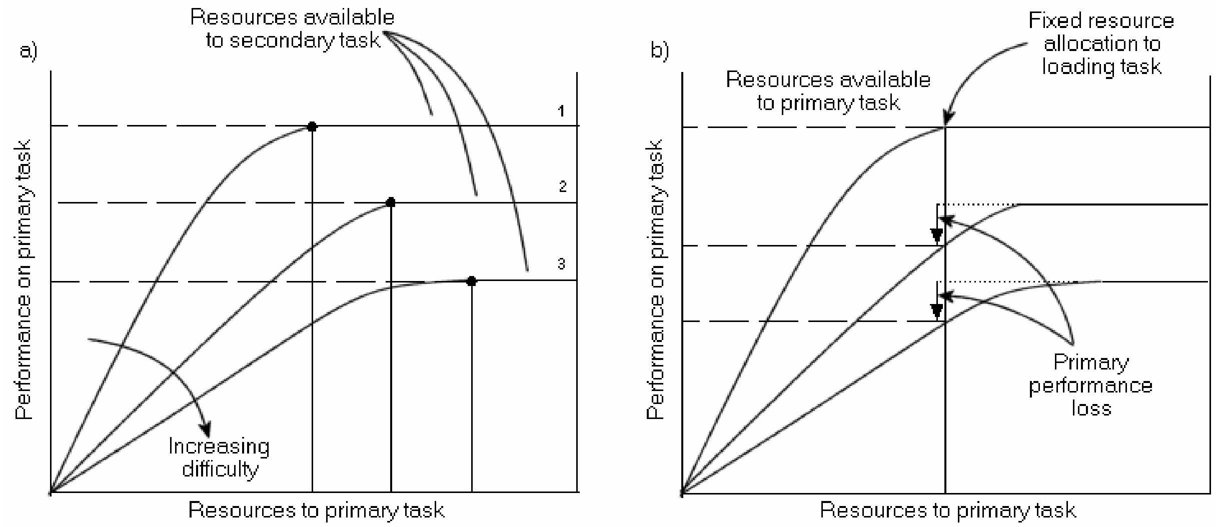

Figure 5. Relationships among the performance-resource function, resource allocation, and primary-task difficulty. (a) Subsidiary secondary task technique, (b) secondary loading-task technique (adapted from Wickens & Hollands, 2000, p. 463).

Figure 5a shows the performance-resource function that has been assumed for the subsidiary secondary task paradigm. Three different primary task performances under different difficulty levels are displayed. As can be seen in the figure, more difficult primary tasks show a decreased maximum performance. Resources available to secondary tasks are variable and dependent on primary task performance. The more difficult the primary task is, the less resources for secondary task performance are available.

As noted before, the same resources have to be used by both tasks and this is also the reason for one of the biggest concerns with this paradigm. A high level of intrusion of the secondary onto the primary task performance is expected if the two tasks use the same resources and have to be time-shared. It could also be that the added task may not only impose additional workload onto the operator, but may fundamentally change the processing of the primary task (Tsang & Vidulich, 2006). Additional problems may occur if the secondary task is omitted due to very high demands in the primary task. Moreover, although operators are instructed to prioritize the primary task, they may follow a different resource allocation policy that can hardly be controlled. Finally, it has been argued that secondary tasks always impose a certain amount of non-specific intrusion in the sense of peripheral interference upon the operator (Eggemeier & Wilson, 1991). Especially for the use in applied environments like car driving it has been pointed out that use of resources can not be completely predicted in advance (De Waard, 1996). For example, although car driving is partly automated and mostly a visual-motor task it is unclear to what extent this task makes use of central processing capacity in a complex traffic environment. In this case, a secondary auditory digit addition task that mostly draws onto the same central resources would then cause an unpredictable interference. Nevertheless, there are numerous studies that successfully applied the subsidiary secondary task even in a real-traffic environment. A recent example is the study by Piechulla et al. (2003) who demonstrated that the varient difficulties of predefined driving situations could be reliably measured by assessing glance frequency towards a continuous secondary visual workload task.

A second approach for the use of secondary tasks is the loading task paradigm. *Figure 5b* illustrates the relationship among the performance-resource function, resource allocation, and primary-task difficulty in the loading task paradigm. Within this paradigm, operators are instructed to completely focus all necessary capacities to the performance of the secondary task. The degree of intrusion of the secondary with the primary task is assessed and the influence of task load onto different primary tasks can be compared. Secondary task performance is measured and it should stay constant across tasks to ensure that the specified criterion levels are maintained. This paradigm is particularly useful when the primary task demand is too low to be sensitive. It is also used to achieve more representative results, i.e. if certain additional demands are expected in an operational environment which would be absent in a laboratory test. As can be seen in *Figure 6*, the addition of the secondary loading task results into a general increase in total workload and therefore to a more sensitive primary task measure.

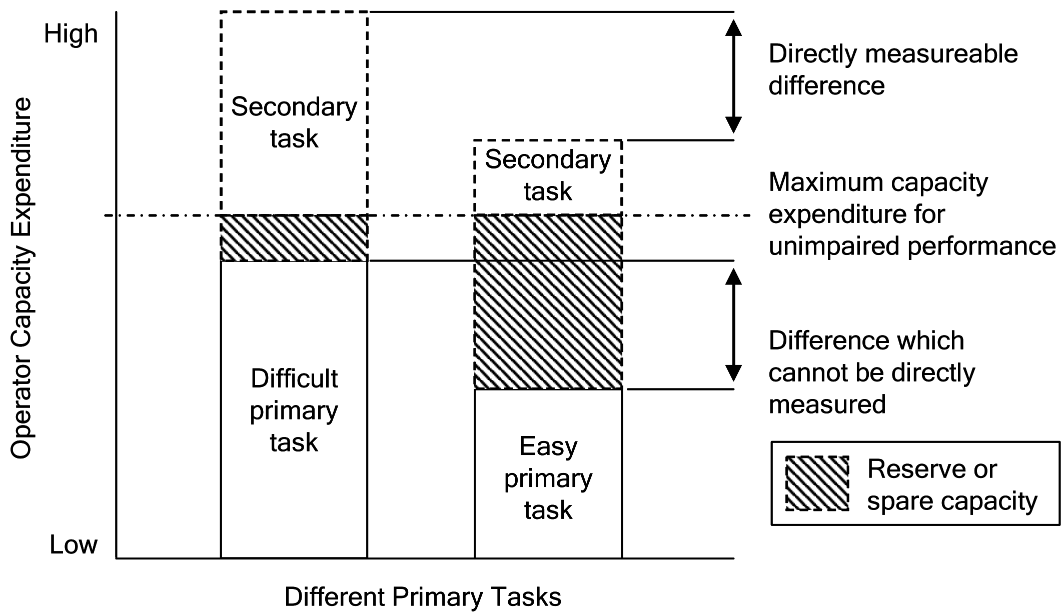

Figure 6. Representation for use of the secondary task to measure operator reserve processing capacity. Neither primary task exceeds operator processing capacity for unimpaired performance. Only by adding a secondary loading task the processing capacity is exceeded and decrements in secondary task performance can be used as a direct resource measure (adapted from O'Donnell & Eggemeier, 1986, p. 42/25).

As pointed out before, a comprehensive knowledge of the type of resources involved in each of the tasks that are under investigation is essential for a correct interpretation of the results. As a possible solution to this problem, it has been suggested that a battery of secondary tasks (e.g. the "criterion task set" proposed by Schlegel, Gilliland, and Schlegel, 1986) might be used, each tapping a different resource. Through a successive reduction of the battery's dimensionality the sensitive resource dimensions for a certain primary task could be identified. Such an approach would allow the construction of a resource/load profile and could lead to a more diagnostic analysis of workload (O'Donnell & Eggemeier, 1986)

To encounter the aforementioned problems of primary task intrusion as well as to avoid unsatisfactory operator acceptance of artificial secondary tasks, it has been proposed that embedded secondary tasks might be used. Embedded tasks are already by definition a legitimate component of the primary task, but not the central component of the task. In this way, it is ensured that the operator does not independently switch priorities between tasks. Moreover, artificial and experimenter-imposed task priorities are avoided. An example of such a task may be the number of rear-view mirror checks when driving is the primary task (e.g. De Waard, 1996).

Finally, another possibility to overcome the problem of primary task intrusion has been introduced as the so-called adaptive task technique. When using this technique, primary task performance is kept at a constant level while secondary task loading is varied to different degrees. The idea is that the level of secondary task loading that can be applied without intrusion represents one measure for primary task workload. For a more comprehensive description and examples please refer to O'Donnell and Eggemeier (1986, p. 42/27).

Several authors listed and evaluated a great variety of secondary tasks and their sensitivity to specific resources and suggested detailed guidelines for their application. For a comprehensive discussion please refer to (O'Donnell & Eggemeier, 1986 or Tsang & Wilson, 1997)

2.3.4 Subjective Workload Assessment Techniques

"No one is able to provide a more accurate judgment with respect to experienced mental load than the person concerned."(De Waard, 1996, p.31) Following this assumption, several questionnaires have been developed to assess mental workload. Only the most popular instruments are presented in the following. Please refer to O'Donnell & Eggemeier (1986) and De Waard (1996) for additional techniques that could not be included here.

A popular one-dimensional questionnaire is the Rating Scale Mental Effort (RSME) developed by Zijlstra and van Doorn (1985). It simply consists of a vertical line from 0 to 150 mm that is subdivided into steps of 10 mm. At several anchor points along the line, statements that reflect invested effort are given. Subjects estimate their own level of mental workload by drawing a cross onto the line. Another similar, but older scale is Bartenwerfer's activation scale (Bartenwerfer, 1969). Its scale ranges from 0 to 270 and different statements are included that aim at the assessment of activation in the sense of a broader, combined measure of mental workload and feelings of arousal.

Especially in the context of driving, video rating has been used to assess the level of mental workload in certain situations (e.g. Schumacher, 2001). The driving situation is recorded by video and displayed to the driver after the drive. The subject is asked to continuously rate the situation by using a sliding lever. In some studies such a lever may even include force-feedback, i.e. sliding the lever becomes more difficult within the extrema of the scale leading to higher accuracy. While this technique holds the advantage of being unobtrusive, it requires very precise instructions for the subject in order to obtain a valid and reliable measurement.

It has often been criticized that one-dimensional scales can not sufficiently cover the multidimensionality of mental workload and it has been argued that an introspective diagnose on such a scale does not allow the tapping of a specific resource demand. However, please note that even with the development of multidimensional questionnaires this problem could not be completely solved.

The most commonly used multidimensional instruments are the NASA Task Load Index (TLX) scale (Hart & Staveland, 1988) and the subjective workload assessment (SWAT) technique (Reid & Nygren, 1988). The NASA TLX measures workload on five 7-point scales, namely mental demand, physical demand, temporal demand, performance, effort, and frustration level. The SWAT scale focuses on three 3-point scales that are time load, mental effort load (conscious mental effort or concentration), and stress load (confusion, risk, frustration, or anxiety). As a result to both multidimensional techniques a final unitary overall score of workload calculation is offered. When comparing both scales, Hill et al. (1992) came to the conclusion that the TLX technique seemed to be more reliable and to provide a higher diagnosticity due to its greater resolution per scale.

In sum, subjective measurement techniques have several advantages. Their application, especially in the case of straightforward and one-dimensional scales, is simple. Moreover, this type of measure is highly immediate. It can provide a measure even before any performance degradation occurs (Muckler & Seven, 1992). However, this immediacy is only possible if the subject is asked within sufficiently brief time intervals. For the use in an applied setting during a continuous and safety critical task like car driving, this is one of the biggest concerns with this technique. Prompting the subject in order to assess mental workload disturbs the driver, directs attention away from the street and introduces additional workload. Although less intrusive methods, e.g. using one-dimensional scales and prompting the driver via touch-screen display have been developed, the level of interference with primary task is still fairly high.

On the other hand, asking the subject after task completion will most likely introduce biases to the measurement like distortions due to personal memory recall.

Finally, as Yeh and Wickens (1988) pointed out, the number of tasks that a participant has to perform strongly influences the subjective rating. Subjects that have to perform two tasks in comparison to one will most likely always report that the workload from the dual-task condition is higher than the one in the single task condition even if the two tasks are not difficult and access separate resources.

2.3.5 Physiological Measures

The last group of workload assessment techniques presented here are physiological measures. The most commonly used physiological measures are: electrocencephalogram (EEG), electrocardiogram (ECG), blood pressure, electrodermal activity (EDA), pupillatory response, eye lid closure frequency, eye movements, endocrinological responses, body temperature, electromyogram (EMG), and respirational indicators. A discussion of each of these measures would be beyond the scope of this work. For detailed reviews of the techniques as well as references to studies that used these techniques please refer to Manzey (1998) or Tsang & Wilson (1997). The spontaneous EEG activity, event-related potentials (ERPs), and the dispersion measures of the ECG were used in the experiments reported below. They are in the focus of sections 2.5 and 2.6 below.

The most important advantages and disadvantages in the use of physiological measures have been summarized by Schmidt (2006). Psychophysiological measures are mostly unobtrusive and the subject's body response is continuously recorded with a high temporal resolution. In this way, the approach is regarded as providing a maximal objectivity since the subject can hardly control the outcome of the recording. With the combined use of multiple techniques highly complex cognitive processes can be reflected in the recordings.

On the other hand, disadvantages of physiological recordings are easily recognizable. A profound knowledge about physiological and psychological processes as well a comprehensive understanding of data analysis methods are important prerequisites for their successful application. Especially the preparation of a multi-channel EEG, but also the setup of other physiological techniques take some considerable amount of time and effort which often leads to a low acceptance of the assessment among subjects.

In sum, while there is a very well established way to physiologically measure physical workload, this is clearly not the case for mental workload. Two reasons are known for this problem: (1) the blurriness of the theoretical concept and (2) a frequently observed divergence of subjective, physiological, and performance measures indicating multiple dimensions of the theoretical concept that has not been comprehensively described yet (Manzey, 1998). As a consequence a multivariate analysis of the different measures has often been demanded (Manzey, 1998). Several authors pointed out that there is an urgent need to go beyond a mere collection of univariate results presented side by side. In this respect, the study of Backs, Ryan, & Wilson (1994) has been mentioned as a positive example. They used principal component analysis to differentiate sympathetic and parasympathetic components out of various data sources including subjective, performance, and physiological measures.

2.4 Mental Workload Assessment During Driving

2.4.1 Advantages of the Secondary Loading Task Paradigm

The topic of mental workload during car driving has been widely studied. Many studies aimed at a better understanding of the driver's current cognitive state during difficult driving situations (e.g. Piechulla et al., 2003, Schmidt, 2006; Verwey, 2000). The authors of these studies chose a straightforward account by assessing mental workload elicited by the driving task in a real-traffic setting. On the other hand, studies investigating workload effects stemming from road or traffic situations always have to overcome the difficulty to reliably implement well controlled conditions of high workload. Especially in real traffic driving, it is very difficult to control all the influences that act upon the driver and potentially cause high workload. Moreover, driving situations of high workload are usually short (e.g. entering a highway) resulting in a limited number of data points, and most of the time, they involve considerably high amounts of muscle activity introducing artifacts to electroencephalographic measurements (see chapter 2.5.5 below for more details on EEG artifacts).

The focus of the present work is slightly different from the approach described above. Its focus lies on drivers' workload elicited by secondary tasks only, while driving has been as much controlled as possible. Moreover, it concentrated on situations in which drivers were mentally occupied by additional tasks that were performed in parallel to a highly automated car driving task. Such an approach holds the advantage to conduct real-traffic experiments in which drivers' mental workload can be selectively manipulated for a sufficiently long enough time to collect a higher number of data points. This approach was used by researchers who tried to address mental workload induced by secondary tasks using various assessment techniques. Most of these works employ either a subsidiary or a loading task (see chapter 2.3.3). However, the analysis of real-traffic driving dynamics in the scope of a secondary loading task design is usually difficult (see Tsang & Wilson, 1997). Neurophysiological measurements such as EEG have the advantage of allowing the investigation of driver workload without interfering with the experimental procedure. EEG studies are rare in driving research on mental workload, especially outside the laboratory. The few existing studies shall be presented in chapter 2.5.3.1 below.

2.4.2 The Investigation of Cell Phone Use During Driving

This section presents a selection of relevant studies employing one or more of the various kinds of assessment techniques. Many valuable insights on mental workload induced by secondary tasks have been gained from the investigation of cellular phone use during driving. On the basis of the collected data existing theoretical accounts have been evaluated and some new perspectives could be contributed. In addition, the authors' experiences on methodological issues can provide valuable assistance in the design of field experiments in a complex real-traffic setting. Although tasks used in the present work were highly artificial and primarily designed for the purpose of measuring the drivers' cognitive state with the use of the EEG, they included some elements that resemble a conversation-like situation inside the car like it has been frequently investigated before. Results may thus be transferable.

2.4.2.1 Contributions from epidemiological and simulation studies.

Caird, Scialfa, Ho, and Smiley (2004) counted 84 epidemiological and driver performance studies covering the period from 1969 to 2004. Up-to-date, only a few studies (e.g. Lissy, Cohen, Park, and Graham, 2000) have argued that the use of cellular phones during driving would be of negligible risk compared to the fact that cellular phones provide numerous benefits to public health and safety. On the other hand, there is empirical evidence speaking for a potential risk caused by mobile phone use.

Two epidemiological studies that pointed out the elevated risk of mobile phone use during driving have been frequently cited. Redelmeier and Tibshirani (1997) analyzed the cellular phone bills of 699 drivers involved in car accidents in the USA. They demonstrated that 170 drivers (24%) were engaged into a cell-phone interaction within the 10-min time period preceding the accident. No significant advantage for the usage of hands-free cell-phones over the usage of handheld cell-phones could be observed which spoke for an interference based on attentional factors rather than interference due to the pure manual manipulation of the phone (e.g. dialing or holding the phone). The authors concluded: "The use of cellular telephones in motor vehicles is associated with a quadrupling of the risk of a collision during the brief period of a call." (p.453). Furthermore, it has been argued that this accident risk is comparable to the increased risk found for driving above the legal US blood alcohol concentration limit at 0.08% weight/volume (see also Strayer, Drews, & Crouch, 2006). However, these data does not automatically imply that the use of cell phones causes an increase in accident risk. For example, it could be that drivers who like to use the phone during driving generally prefer a riskier driving style (Wilson, Fang, Wiggins, & Cooper, 2003). Moreover, a highly emotional state may cause erratic driving as well as it may increase the probability of talking on the phone.

An Australian case-crossover study by McEvoy and colleagues (2005) evaluated the cellular-phone records from 456 drivers who needed hospital attendance after an accident and who were on the cell-phone 10 min before the motor-vehicle crash. They replicated the results of Redelmeier and Tibshirani (1997) and they pointed out that even with voice activation and thus totally hands-free phone use, the accident risk due to the profound effects on drivers' attentional capacities still remains.

Many experimental studies have been conducted to find an adequate description of the cognitive processes underlying driver workload and distraction. Strayer and Drews (2007) reported four driving simulator studies that they executed to assess the degree of distraction of cell-phone use on car driving. The authors referred to "inattentional blindness" to describe the effect that they observed, i.e. even if the drivers detected relevant objects like traffic signals, they were less likely to create a longer lasting memory of those objects. By the use of event-related potentials (P300, see chapter 2.5.4.1 for more details on this component), the authors pin down the source of the performance decline to be related to false encoding of the detected objects. They argued that Wickens' multiple resource model (1984) does not apply in the case of cell phone conversations. They postulate that interference problems would rather be related to a central processing bottleneck (Welford, 1952, for review see Pashler & Johnston, 1998) that forces serial processing of the driving task and conversation-related information. Conversing on the phone is especially prone to interference in this case, because it is a continuous task that is composed of "turns", i.e. one person is constantly talking until it is the other person's turn to talk. Unlike in a live conversation, the two conversation partners do not share the same sensation from the outside world when they talk on the phone. The task can not be easily interrupted or forced to follow a different processing mode than that of taking turns. Strayer and Drews concluded that the demands stemming from conversations should be qualitatively different from other auditory, verbal, or vocal tasks that are commonly performed in parallel to the driving task.

In line with these results, an earlier study by Strayer and Johnston (2001) showed that the level of driving impairment was similar between drivers who talked over a hands-free or a handheld phone. Two experiments employed different cognitively loading tasks and assessed their degree of interference onto braking and lane keeping performance. No impairment could be observed for the conditions in which drivers were simultaneously listening to the radio or a book on tape as well as when they performed a continuous shadowing task using a handheld phone. By employing a post-test, the authors ensured that drivers were indeed listening to the book on tape. Results showed that only the conditions in which drivers were engaged into a conversation or performed a word generation task led to significant performance attenuations. As in Strayer and Drews article from 2007, the authors claimed that their results provided evidence against Wickens' multiple resource model. The shadowing and word generation task should address similar stages of processing, i.e. a similar degree of overlap in resources between primary and secondary tasks should be provided. The only difference between these tasks was the level of attentional demand imposed by the generation of words. However, against the predictions of the multiple resources model, only the word generation task interfered with the primary task.

Levy, Pashler, and Boer (2006) demonstrated that the central-bottleneck theory held true for braking in a car-following task performed in a driving simulator in parallel to different types of secondary stimulus-response tasks (auditory vs. visual and manual vs. vocal response). Longer braking reaction times were observed when the stimulus onset asynchrony between the two simultaneous tasks was decreased. The results are in accordance to the theory and speak for the fact that vehicle braking is affected by a central bottleneck. Their work demonstrated that response modality (vocal or manual) from a concurrent task did not differentially impair driving, but that even if two tasks do not overlap in their modality-specific resources interference between tasks can occur.

Besides the results from the simulator experiments above, other experiments (for a review see Goodman, Tijerina, Bents, & Wierwille, 1999) observed significant interference between phone use and driving. As early as 1969, Brown, Tickner, and Simmonds demonstrated that the use of a radiophone severely interfered with the drivers' decision making processes while automated driving tasks were only minimally impaired. In the following years, numerous studies followed showing significant influences on reaction times (Consiglio, Driscoll, Witte, & Berg, 2003; Strayer, Drews, & Johnston, 2003; Alm & Nilsson, 1994, 1995), variability of lane position and speed (Alm & Nilsson, 1994), following distance (Alm & Nilsson, 1995), and situational awareness (McKnight & McKnight, 1993; Strayer et al., 2003). In a recent study, Beede & Kass (2006) showed that simulated driving performance assessed via number of traffic violations (e.g. speeding), attentional lapses (e.g. stops at green lights), driving maintenance (standard deviation of lane position), and response time (braking reaction time) was significantly impaired by a cellular phone conversation that involved visuo-spatial content and that peripheral visual signals were more often overlooked.

There is a significant lack of studies investigating electroencephalographic correlates of high mental workload induced by distracting secondary tasks during driving in real traffic. The presented studies address important theoretical aspects in trying to describe the cognitive mechanisms underlying mental workload. Conclusions on existing theoretical models are transferable to the research question at hand. Based on the empirical data, it can be concluded that secondary tasks like a phone conversation are highly distracting and may have severe consequences for the drivers' ability to simultaneously perform another task. Existing theoretical accounts predicting only a slight or no interference from an auditory-verbal secondary task onto drivers' mental workload have been seriously questioned. Some authors postulated that the observed constraints imposed upon the driver could be better described in terms of a central processing bottleneck rather than in terms of multiple overlapping resources. The central processing bottleneck may be the cause for errors in encoding incoming information leading to inattentional blindness.

However, the short review above straightened out a second argument. It demonstrated once more the limits of secondary task paradigms. The interpretation of human performance data always implies strict theoretical assumptions. Within these paradigms mental workload is not directly assessed, but it is inferred from performance data and other data sources. Although laboratory studies allow a maximum of control, the need for an unobtrusive and direct measurement technique becomes evident. This is especially true in the light of complex driving field experiments as those reported in the following.

2.4.2.2 Evidence from real-world driving studies.

It is well known, that task involvement in simulator environments is generally low and it has been doubted that results would be applicable to drivers in real traffic (McEvoy et al. 2005, Hole, 2007). A large number of on-road studies (e.g. Angell et al. 2006; Brookhuis, De Vries, & De Waard, 1991; Cooper, Zheng et al., 2003; Harbluk, Noy, Trbovich, & Eizenman, 2007; Jahn, Oehme, Krems, & Gelau, 2005; Lamble, Kauranen, Laasko, & Summala, 1999; Nunes & Recarte, 2002; Patten, Kirchner, Östlund, & Nilsson, 2004; Recarte & Nunes, 2000, 2002, 2003; Recarte, Nunes, & Conchillo, n.d.) have been published that investigated driver mental workload induced by secondary tasks. A detailed review of these on-road studies is beyond the scope of this work. Several authors (e.g. Green 2004; Hole, 2007) provided detailed study reviews. In sum, results from on-road studies confirmed the findings from the laboratory experiments described above.

The works of Nunes, Recarte and Conchilla (Recarte & Nunes, 2000, 2003; Recarte et al., nd. cited after Recarte & Nunes, 2003) provided important methodological impulses for the design of the on-road driving experiment in this work, i.e. a secondary loading task and a tertiary workload metrics task were used in Experiment 2. Recarte and Nunes (2003) ran a real traffic driving experiment in which drivers were confronted with several secondary loading tasks and a tertiary peripheral detection task. Pupil size and spatial gaze behavior were recorded with an eye tracking system and the visual detection performance from the tertiary task was evaluated. To induce mental workload the drivers had to perform eight different tasks: a production vs. an acquisition task with abstract or concrete content and two different versions of daily life cognitive tasks performed on the phone or in live conversation correspondingly. Their data demonstrated that mentally loaded drivers produced spatial gaze concentrations and showed impairment in visual-detection which was qualitatively different from a tunnel vision (Rantanen & Goldberg, 1999). They claimed that their results reflect endogenous distraction, i.e. the results transferred to the situation when drivers were lost in their own thoughts or engaged in cognitive activity unrelated to the driving task. This was opposed to exogenous distraction which captures attention and gaze of the driver due to external objects or events (Posner, 1980). Most interestingly and in line with laboratory results reported above, only production tasks and complex conversations indicated danger for road safety. In contrast, verbal acquisition tasks were harmless and did not interfere with safety relevant performance. However, these tasks were also rated as easy by the participants, thus the reason for the obtained null effect may have lain in a low mental workload manipulation rather than the characteristics of the task. Different from Strayer et al. (2001, 2007), Recarte and Nunes (see also Nunes & Recarte, 2002) did not see a difference between a phone and a passenger conversation. They argue that both types of conversation would be equally dangerous for safe driving and that "the complexity of the message is what matters" (p.135).

In an earlier study, Recarte and Nunes (2000) already provided some on-road evidence for the impact of endogenously distracting mental tasks. They observed a significant increase in pupil size, a spatial gaze concentration, and a reduced inspection frequency of mirrors and speedometer when drivers were under high workload. In contrast to other simulation (e.g. Alm & Nilsson, 1994) and on-road studies (e.g. Brookhuis, De Vries, & De Waard, 1991) driving performance measures remained unaffected in their experiment. High spatial imagery content of the loading tasks resulted in longer fixation durations and more pronounced workload effects.

Recarte et al. (n.d.; cited after Recarte & Nunes, 2003) compared production tasks (continuous verbal responses) and acquisition tasks (attending to and memorizing audio messages) with either verbal or spatial content by assessing their influence on visual search behavior. They found that either type of production task had a significant influence, but that acquisition tasks did not affect visual search behavior at all. However, as discussed by the authors it remained unclear how these changes in visual search behavior related to the probability of missing relevant information or making wrong decisions, i.e. the optimal spatial gaze distribution for drivers' maximum situational awareness was unknown. Up-to-date, there is no "gold standard" for visual search behavior to compare their results to.

The studies described above give an idea on how researchers have conducted their investigations outside the laboratory. Moreover, the presented collection of on-road driving experiments demonstrates that a well-controlled assessment of mental workload is feasible in the field and that basic conclusions drawn from simulation data may be applied to real-traffic situations. Despite the high amount of effort in preparation and running of such experiments, the advantages are worth the costs.

The results of on-road studies possess high face validity and they can provide valuable insights for promising future applications.

2.4.3 The Development of Workload Managers

A considerable amount of on-road studies on mental workload and driver distraction are studies concerned with the development of workload managers. In the light of these studies, the high benefit that an EEG measurement could provide becomes evident. The EEG is fast, direct and it involves minimal interference with the performed task. These advantages could make it an important tool for the development and evaluation of workload management systems. “A workload manager is a system that attempts to determine, if a driver is overloaded or distracted, and if so, alters the availability of telematics and the operation of warning systems.” (Green, 2004, p.1). The aim of such a workload manager is to assist the driver in adequately coping with potentially dangerous situations of high mental workload and to reduce the risk of an accident.

Green (2004) presented a coarse categorization of workload management systems of four different groups depending on what is measured: (1) driving situation, (2) driver input, (3) vehicle performance, and (4) driver state. For example, workload managers assessing the driving situation usually involve so-called “digital maps”, which contain information about the driving situation, e.g. the type of road a driver is on or information about usual traffic density. Most systems that have been developed combine these various sources of data. According to Green the implementation of such a system may decide whether a warning system is useful or not, i.e. workload managers may help to avoid false alarms. A workload manager would be able to sense driver inattention to the road and would only intervene in situations of which the driver is really unaware. Reducing the number of false alarms helps to increase operator's acceptance. The system is generally perceived as more reliable and drivers will trust it more. Besides the fine-tuning of existing warning systems, a workload manager may also simply be used to protect the driver from too much incoming information or help using scarce cognitive resources most efficiently. For example, incoming phone calls can be directed to a voice mail system if the driver is in a state of high workload. Moreover, messages could be prioritized and depending on the current attentional state of the driver they may only be immediately presented if absolutely necessary. Some manufacturers (e.g. Mercedes-Benz) already use navigation systems that automatically switch from auditory to visual information display if the driver is receiving an incoming phone call. Similarly, driver assistance systems may adjust the modality of warning messages to suit the driver's current attentional state.

Piechulla (Piechulla, 2006; Piechulla et al., 2003) described a real-traffic experiment in which a workload management system was assessed by using subjective, behavioral, and multiple psychophysiological measures. Information about road difficulty based on Fastenmeier's (1995) driving situation classification system and information about current driving dynamics were used to identify high workload situations on a predefined route. The data was used to trigger two different types of driving assistance systems to ease the mental workload of the driver. In the experiment, 12 drivers (6 novices, 6 experienced) either drove a baseline drive without assistance, an assisted drive including adaptive cruise control and camera-based steering support or an assisted drive with adaptive cruise control, steering support and automatic suppression of incoming phone calls during workload critical situations. Mental workload was measured by assessing heart rate, heart rate variability (0.1 Hz component, see chapter 2.6.2), non-specific skin conductance response, lateral frontalis-electromyogram, glance frequency during the performance of a visual task with a scrolling text display, and a questionnaire. Subjective measures taken after each completed drive failed to reveal any difference in driving conditions. The authors observed that psychophysiological measures showed tendencies which spoke for the sensitivity of the assessment techniques. Only secondary task performance assessed with an eye-tracker revealed statistically significant results. The data showed a reduced frequency of glances to the in-car display while the drivers were engaged into higher demanding driving situations. Piechulla (2006) concluded that in the scope of adaptive workload management systems, eye tracking systems would be the most promising measures since they directly assess drivers' visual attention to the road. The study showed the importance of a reliable and instantaneous real-time assessment of mental workload for the development of workload management systems. A prerequisite for such an assessment is to adequately understand the complexity of human behavior in real traffic.

The list of EU and US projects on workload management systems is long and despite the fact that many Original Equipment Manufacturers (OEMs) and suppliers put a lot of effort in the development of such a system, no comprehensive solution has been found so far (see Green, 2004). Up-to-date only Saab (model 9-3 and 9-5 built after 2003) and Volvo (Intelligent Driver Information System, S40 and V50; Volvo, 2004) offer simple versions of dialog managers for their customers (Green, 2004; Piechulla, 2006). Workload managers are still in a state of development and many car manufacturers have hesitated to take them to series production. One of the major problems still lies in the accurate prediction of dangerous cognitive states. A workload manager that does not react until a driver is already in a state of high workload is barely useful. Another issue concerns the reliability of detection which is directly connected to the operators' acceptance of the system. Finally, situations of high workload are usually rare, e.g. consider the probability of receiving a phone call while passing a car on a two lane street. The frequency of an intervention by the workload manager would be hardly noticeable. However, the long list of past and ongoing research projects shows that workload managers have been perceived as holding the potential to provide a valuable contribution to increased road safety in the future.

2.5 The EEG Measurement Technique

Richard Caton (1875) was among the first who reported about electroencephalographic (EEG) recordings from monkeys and rabbits. It took over 50 years, until in 1929, Hans Berger succeeded in reliably recording the EEG from a human scalp. He established a first categorization of brain oscillations into alpha (8 - 13 Hz) and beta waves (14 - 30 Hz). Ever since then the EEG has found a wide application in determining brain states in a variety of cases, e.g. in diagnosing pathological conditions or in sleep research. Two different types of EEG recordings have been established: (1) the spontaneous activity or (2) the activity time-locked to a stimulation known as event-related potentials or evoked potentials. Compared to other recording techniques the EEG holds an important advantage: speed. Neural activity is recorded at a very high time resolution, although only a coarse localization of the activity’s neural generators in the brain is possible. Most recently, the EEG has been combined with functional neuroimaging techniques to attain an answer to both where and when brain activity occurs (Bösel, 1996; Winn, 2001).

2.5.1 Origins of the EEG Signal

EEG is measured using scalp electrodes which record the difference in the electric potential between an electrode with an active neural signal and an electrode placed over a supposedly inactive region that serves as a reference. The measured potential usually varies between 0 and 200 μV (Schandry, 2006). Hebb (1949) postulated that information in the brain is always represented by groups of neurons, so-called cell assemblies, which are jointly active. The extent and structure of those cell assemblies may change over time.

Due to the cytoarchitectural structure of the neocortex, cell assemblies elicit a combined dipole in the electrical field over the scalp. The field potentials are mostly composed of excitatory post synaptic signals which arrive at the apical dendrites in layer I and II of the neocortex and which originate from the thalamic nuclei, commissural and long association fibers. EEG signals represent the summated electrical activity of the brain generated by a variety of neural structures and transmitter systems. Inhibitory and action potentials only show a relatively low contribution (Birbaumer & Schmidt, 2006).

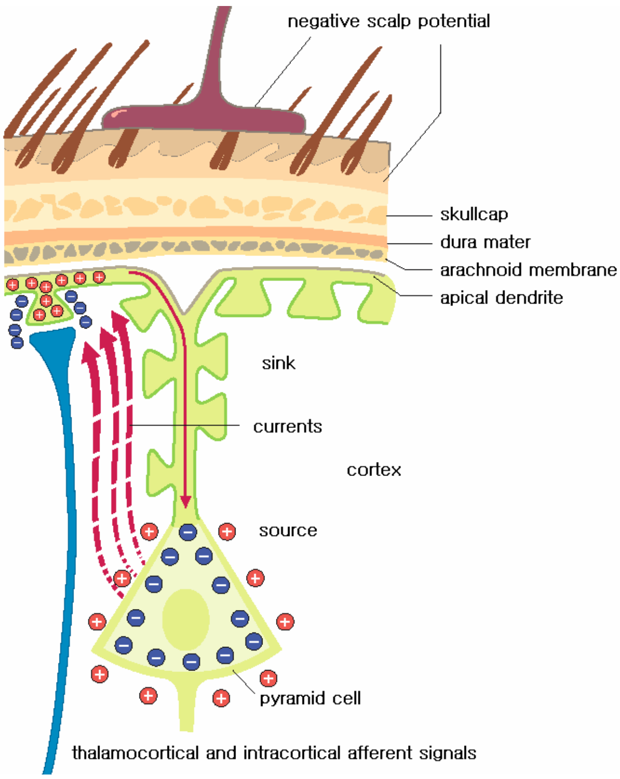

Figure 7. A cortical dipole (adapted from Birbaumer & Schmidt, 2006, p. 472)

Figure 7 illustrates the generation of a cortical dipole after the arrival of an afferent signal from the thalamus or other cortical regions. Due to the incoming flow of Na+ ions at the apical dendrites and the resulting flow of current (at the sink) towards the cell body, the extracellular regions around the dendrites become negatively charged and a negative field potential is created. The lower layers around the cell body are positively charged and an extracellular voltage exchange is initiated from the cell body (the source) towards the regions of high depolarization at the apical dendrites. Thus, the negatively charged upper layers and the positively charged lower layers constitute the dipole at the cerebral cortex.

Hebb regarded the synchronization of oscillating neural discharges as the underlying mechanism for the ever changing activation of temporary neural coalitions. So-called reverberatory circuits which can be held active after the stop of input are expected to be the prerequisite for neural plasticity and learning. Today, it is well established that recurrent fiber loops are responsible for the oscillating neural discharges. Based on analysis techniques like frequency analysis, the cognitive functionalities of these oscillations have been investigated.

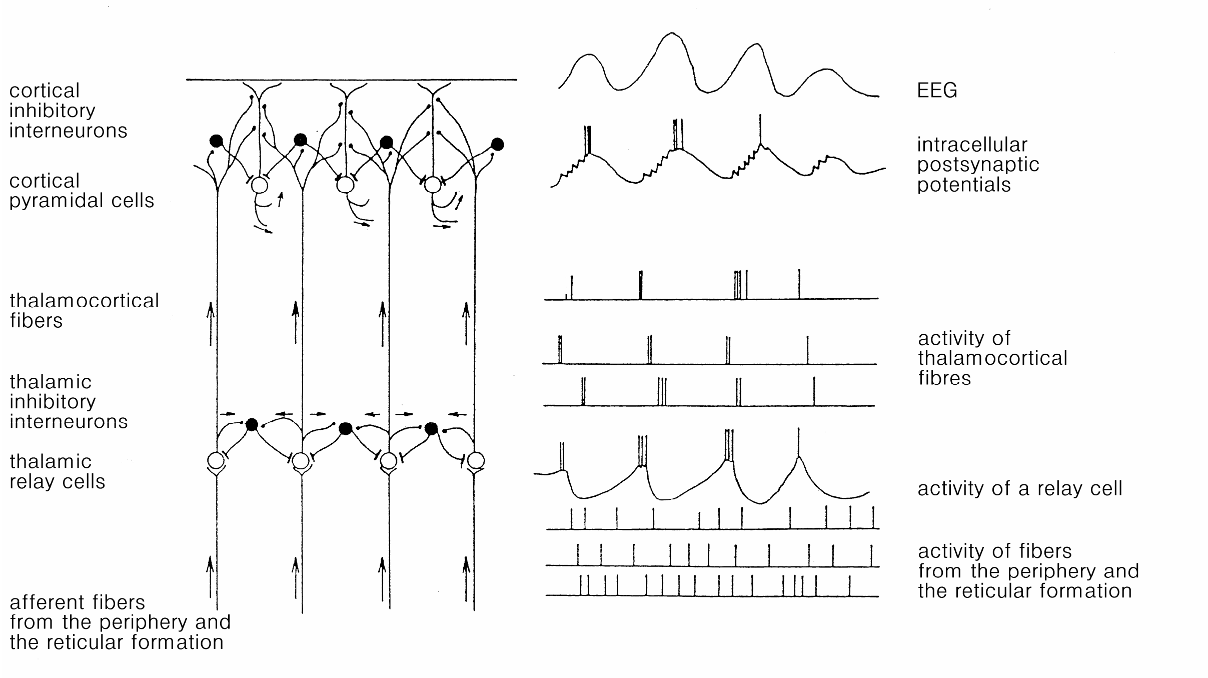

Figure 8. The generation of alpha oscillations in the EEG (adapted from Bösel, 1996, p. 19)

The generation of the most prominent brain rhythm, the alpha rhythm, has been extensively described. *Figure 8* illustrates the generation of alpha rhythm in the EEG by showing cortical structures involved and their associated electrical activity. The alpha rhythm is the result of a complex interplay of information from the reticular nucleus, the thalamus and inhibitory inter-neurons. The cells of the thalamus are activated by irregular impulses. After exceeding a certain threshold, reverberatory circuits in the thalamus produce an oscillatory cell activity. This leads to the aforementioned phasic excitatory synapse potentials at the apical dendrites. The summation of these potentials causes a rhythmic change in the electrical field (Bösel, 1996).

2.5.2 EEG Measures

Three basic types of analysis strategies are commonly used for EEG data (Bösel, 1996): (1) the evaluation of the raw EEG (e.g. for diagnosis of epilepsy or sleep research), (2) the analysis of event-related potentials (ERP), or (3) the power spectrum analysis. The latter two methods have found application in experimental research on mental workload and shall be briefly outlined.

2.5.2.1 Event-related potentials (ERPs).

The basic principle of the ERP analysis is that a certain event (external stimulation or cognitive operation) elicits a characteristic brain response which is reliably measurable over several repetitions. The magnitude of the ERP signal is small (5 - 10 μV) in comparison to the amplitude of the background EEG (0 - 200 μV) from which it has to be extracted. Although there has been some effort to evaluate potentials based on single trial events, the classic approach is to average the signal over several incidents (> 20 trials) to obtain a stable response with a sufficient signal-to-noise ratio.

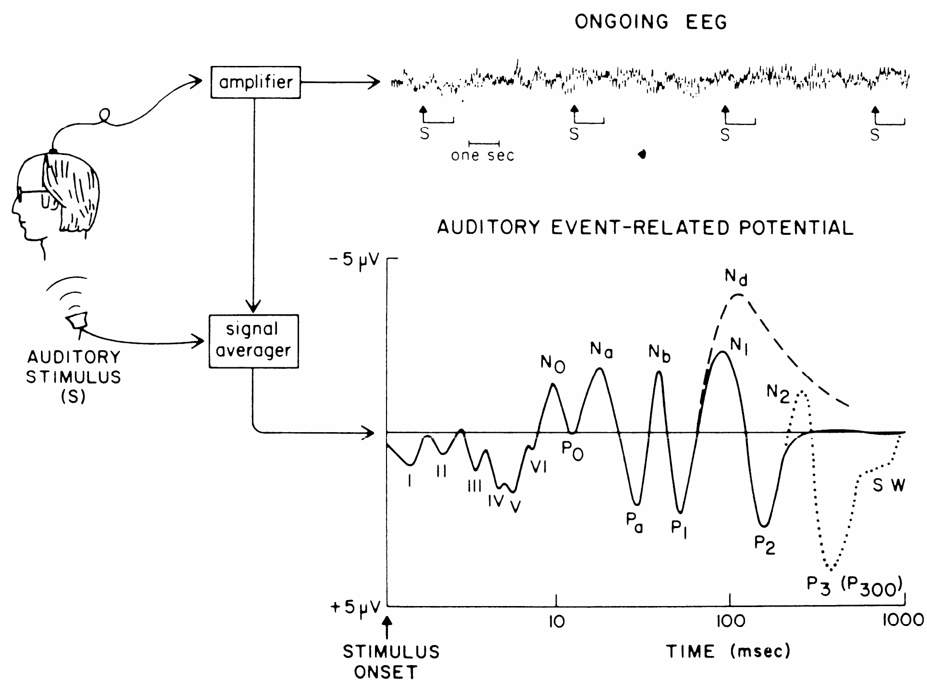

Figure 9. Idealized auditory ERP waveform (adapted from Coles and Rugg, 1997, p.6)

An idealized ERP waveform elicited by an auditory stimulus is shown in *Figure 9*. The figure displays a number of positive and negative voltage peaks (components) that have a specific latency and distribution over the scalp. ERP components are usually given labels that refer to their polarity and position within the waveform (e.g. P1 or N1) or a number that is marking their latency onset (e.g. N100, P300). In other occasions the labels denote a functional description (e.g. readiness potential, RP) or refer to the brain area that is the presumed neural generator of the component (e.g. auditory brainstem response, ABR). The simple labeling by polarity and position may sometimes lead to confusion (Luck, 2005): sensory components from different modalities (e.g. P1 or N1 from the auditory or visual modality), but with the same label usually don't necessarily share the same functional meaning. They just coincidentally share the same polarity and ordinal position in the waveform.

Two classes of components have been distinguished (e.g. Luck, 2005): exogenous and endogenous components. The characteristics (amplitude, latency, and distribution) of exogenous components (e.g. C1) seem to rely on the physical properties of sensory stimuli, i.e. their modality and intensity. It has been assumed that these components are independent from the subjects' state or cognitive interaction with the stimulus. On the other hand, endogenous components (e.g. P300, N400) seem to change as a function of factors like attention, task relevance, and task demand. These components may even arise in the absence of an external event, i.e. in the case that an expected stimulus is not displayed. However, as Coles and Rugg (1997) pointed out, this dichotomy is a coarse oversimplification: "Almost all the early 'sensory' components have been shown to be modifiable by cognitive manipulations (e.g. attention) and many of the later 'cognitive' components have been shown to be influenced by the physical attributes of the eliciting conditions (e.g. modality of the stimulus). For this reason it appears to be more accurate to conceive of an exogenous-endogenous dimension that is roughly coextensive with time. Thus, those ERP components that occur within the first 100 ms of stimulus presentation tend to be more exogenous, while those occurring later tend to be more endogenous." (p. 16). In the present work, early and late components have been analyzed with respect to mental workload. Studies that used similar components to study attention and the availability of mental resources shall be reviewed in section 2.5.4 below. A comprehensive review of ERP components and their functional meanings can be found in the books by Rugg and Coles (1997) and Luck (2005).

2.5.2.2 EEG oscillations.

The oscillatory activity of the spontaneous EEG has been categorized into five different frequency bands that are shown in *Table 2*.

Table 2. The five EEG frequency bands.

Frequency band (Type of waveform)	Frequency range in Hz	Amplitude range in µV
Delta (δ)	0.5-4	20-200
Theta (ϑ)	5-7	5-100
Alpha (α)	8-13	5-100
Beta (β)	14-30	2-20
Gamma (γ)	30-100, mostly around 40 Hz	2-10

Note: Adapted from Schandry (2006, p. 565).

Figure 10 shows the four "classic" frequency bands delta to beta which are visible to the naked eye in the continuous EEG. The gamma-band is generally only observable after specific stimulation (Schandry, 2006).

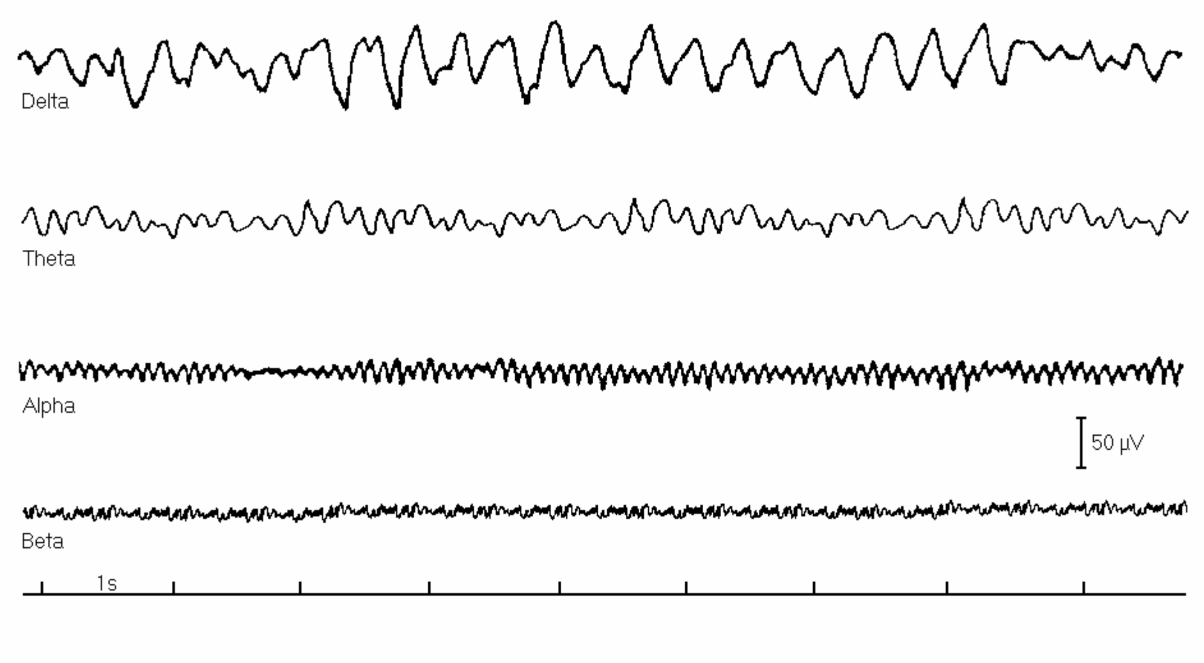

Figure 10. Examples of the most common oscillations in the continuous EEG (adapted from Schandry, 2006, p. 566).

The frequency bands have been related to different cognitive functions. A global categorization has been made by classifying the power bands according to a person's level of activation, i.e. apart from a few exceptions it can be said that the less activated the cognitive state of a person is, the slower the EEG and the higher the amplitude (Schandry, 2006). The lowest level of activation has been related to the EEG delta band (0.5 - 4 Hz). It is well known from sleep research that delta band activity represents the onset of deep sleep phases in healthy adults. In addition, contaminating eye activity is mostly represented in the EEG delta frequency band (see section 2.5.5 on artifacts below).

The theta band (5 - 7 Hz) has been observed at the transition stage between wake and sleep state. Some works reported theta activity under deep relaxation or meditation (e.g. Hebert & Lehmann, 1977; Kubota et al., 2001). Generally, theta seems to be involved in high concentration, e.g. the central executive function in working memory (e.g. Schack, Klimesch, & Sauseng, 2005), especially in encoding and retrieval (Ward, 2003). Theta has also been related to human spatial cognition (Kahana, Sekuler, Caplan, Kirschen, & Madsen, 1999) and it has been shown to increase with higher cognitive task demand (e.g. Gundel & Wilson, 1992; Gevins, Smith, McEvoy, & Yu, 1997). Kubota and colleagues argued that frontal midline (fm) theta occurs with the activation of attentional systems in prefrontal circuitry that involves the anterior cingulate cortex.

Alpha band (8 - 13 Hz) activity is easily visible to the naked eye due to its high amplitude and regular oscillations with a maximum over parietal and occipital electrodes in the continuous EEG. High alpha power has been primarily known to reflect a state of relaxed wakefulness with low visual attention. The brief attenuation (0.5 to several seconds) of alpha power with eye opening is a stable phenomenon named "alpha blockade" (Schandry, 2006). However, recent results suggested that alpha does not necessarily only represent cortical idling, but that it is involved in auditory attention processes and inhibition of task irrelevant areas to enhance signal-to-noise ratio (Cooper et al., 2006; Cooper, Croft et al., 2003; Jensen et al., 2002; Klimesch, Sauseng, & Hanslmayr, 2007; Ward, 2003). Moreover, some researchers divided the alpha band into sub-bands to achieve a finer grained description of its functionality (e.g. Klimesch, 1999). Consolidated findings have been compiled in the following sections below (see section 2.5.3).

Alpha and beta waves are the predominant brain rhythms that are usually observed in a state when someone is awake (Schandry, 2006). The cognitive activity associated with a high power in the beta band has been described as a concentrated state of being awake. It has been regarded to mark visual concentration and the orienting of attention (Birbaumer & Schmidt, 1996). According to Bösel (1996) several beta sub-bands have been differentiated. The beta 1a band (12.5 - 14.5 Hz) has been linked to the inhibition of phasic movements during sleep (Holcombe, Sterman, Goodman, &

Fairchild, 1979; Sterman, 1981). The beta 2 band (> 18 Hz) has been related to the dopaminergic system (Bouyer, Montaron, Fabre-Thorpe, & Rougeul, 1985; Gruzelier, Liddiard, Davies, & Wilson, 1990).

The gamma band (> 30 Hz) has been in the focus of many recent investigations that studied cognitive functions like visual perception, attention, learning, and memory (for review see Jensen, Kaiser, & Lachaux, 2007 or Kaiser & Lutzenberger, 2005). A great number of experiments studied the representation and analysis of perceived objects in the human brain. Today, it is well established that gamma activity is involved in the binding phenomenon, i.e. the coupling of adjacent neural populations resulting in synchronous activity in order to achieve a coherent cognitive representation of an object, instead of a mere sum of singular features of color, shape, or movement (e.g. Keil, Müller, Ray, Gruber, Elbert, 1999). Tallon-Baudry and colleagues (Tallon-Baudry, Bertrand, Hénaff, Isnard, & Fischer 2005) reported about two cortical areas (lateral occipital cortex and fusiform gyrus) known to play a key role in visual stimulus encoding and reliably showing large gamma oscillations that were differently affected by attentional modulation. Sederberg et al. (2007) presented intracranial recording data from epilepsy patients and demonstrated that an increase in gamma oscillations was related to successful verbal memory formation. Finally, a link has been demonstrated between gamma-band activity and neuropsychiatric disorders like schizophrenia, Alzheimer’s disease, and attention deficit hyperactivity disorder (ADHD) (Herrmann & Demiralp, 2005). The authors speculated about a gamma axis of neuropsychiatric disorders.

Besides the classic EEG rhythms further waveforms have been observed. For example the μ-waves or sensomotoric rhythm (SMR) which lie in an almost similar range like the alpha waves (10 - 15 Hz), but occur over central rather than occipital sites. It has been assumed that the μ-rhythm is neurophysiologically identical to sleep spindles and it is known that its power decreases with movements or the imagination of movements (Birbaumer & Schmidt, 2006).

The frequency limits of each band have been arbitrarily chosen and sometimes they may vary to a certain degree between authors. Some investigators declared that it would be necessary to determine frequency bands via a post-hoc factor analysis for each investigation (e.g Schimke, Klimesch, & Pfurtscheller, 1990) or that an individual adjustment of the frequency band limits would be required (e.g. Klimesch, 1999). For some frequency bands, e.g. the gamma band (30-100 Hz) a broad frequency range has been defined and future research might lead to a finer grained definition of sub-bands that are related to specific cognitive functions (Jensen et al., 2007).

2.5.3 The Alpha Band as a Measure of Mental Workload

This section describes research efforts to investigate how human alpha oscillations relate to mental workload and attention. A great amount of investigations concentrated on mental state classification in an applied context and a selection of relevant studies will be presented in the first part of this section. Although these works achieved intriguing classification results, mostly based on more than one data source, some questions related to the electrophysiological underpinnings of mental workload had to be left unanswered. Results from basic laboratory EEG research can provide complementary insights for a more profound understanding of cognitive function in complex multiple-task environments. Therefore, part two of this section has been dedicated to a compilation of these studies' outcomes.

2.5.3.1 EEG research in applied settings.

Up-to-date no real-traffic EEG study has been published that investigated the drivers' mental workload induced by secondary tasks and there has not been enough research to allow a clear statement about the effects of mental workload imposed by secondary tasks onto the drivers' EEG alpha band. Generally, research using EEG alpha power to study neurophysiological correlates of mental workload during driving is rare. On-road and driving simulator EEG studies mostly focused on states of low vigilance (Brookhuis & DeWaard, 1993; Cerezuela, Tejero, Chóliz, Chisvert, & Monteagudo, 2004; De Waard & Brookhuis, 1991; Schier, 2000; Tejero & Choliz, 2002; Zeier & Bättig, 1977), drug influence (Brookhuis, Louwerens, & O'Hanlon, 1986) or sleepiness and fatigue (Kecklund & Åkerstedt, 1993; Lemke, 1982; Lin et al., 2005; Miller, 1995; Wijesuriya, Tran, & Craig, 2007; for review Lal & Craig, 2001a, 2001b; for countermeasures see: Gimeno, Cerezuela, & Montaés, 2006; Lal, Craig, Boord, Kirkup, & Nguyen, 2003). Theses studies demonstrated that an increase in alpha power or an increase in the relative energy parameter $[(\theta+\alpha)/\beta]$ (e.g. Brookhuis & DeWaard, 1993) may be used as an index for a driver's attentional state.

So far driver mental workload has only been assessed by one on-road EEG study (Schmidt, 2006). In contrast to this research, the present work used a secondary loading task paradigm to impose mental workload onto the driver, while Schmidt focussed on the detection of workload elicited by the driving situation. He ran a real-traffic experiment with 20 subjects to identify psychophysiological correlates of mental workload in highly demanding driving situations by using EEG, ECG, and EMG. As previous studies (e.g. Piechulla et al., 2003), he used Fastenmeier's (1995) classification system to identify driving situations of different difficulty levels on a predefined route. Schmidt showed that delta, theta, and alpha frequency bands of the EEG were decreased while beta and gamma bands were increased with a higher difficulty of the driving. Moreover, heart rate and neck muscle activity were significantly amplified in those situations. As Schmidt stated, a potential impact of muscle artifacts on the data could not be excluded, since difficult driving situations like entering a highway were almost always linked to corresponding driver movements.

EEG alpha correlates of mental workload in dual-task situations have been more widely assessed in applied contexts other than driving. Berka et al. (2007) provided a summary work from various research groups that pursued the same overall goal: to use psychophysiological measurement techniques including EEG to develop brain-based adaptive systems that can be employed in a complex real-world environment in order to achieve more accurate and efficient methods for humans to interact with technology. Laboratory works from 1980 - 2000 concentrated on identifying spectral components of the EEG that characterize a state of high mental workload. Numerous works demonstrated that with increased cognitive processing effort alpha waves decrease (e.g. Gevins & Smith, 2003; Glass, 1966; Gundel & Wilson, 1992) and theta activity is enhanced (e.g. Rugg & Dickens, 1982; Mecklinger, Kramer, & Strayer, 1992). Although representing an oversimplification of the complex interplay of cortical networks (see section 2.5.3.2), these results paved the way for some powerful applications. Pope and colleagues (Pope, Bogart, & Bartolome, 1995) were among the first to show a working adaptive automation system which used the degree of operator engagement to control the level of task automation. The system was based on an index of engagement based on the power from the theta (4 - 8 Hz), alpha (8 - 13 Hz) and beta (13 - 22 Hz) bandwidths.

Subsequent works equally succeeded in presenting efficient mental state classification of data from complex environments. The used classification algorithms were based on a bottom-up, data-driven analysis, i.e. most of these works included more or less a mixture of all EEG bands, all electrode positions (Berka et al., 2007, Wilson & Fisher, 1995), and additional data sources like the EOG, ECG, respiration, or eye activity (Ryu & Myung, 2005; Wilson, 2002). By the use of multivariate methods and complex classification rhythms, reasonable classification accuracies could be achieved. On the other hand, theses works showed that the EEG by itself was not sufficient to provide all the information needed for an adequate workload classification. Although "EEG measures may be more sensitive to task variables than either performance or subjective measures or the other physiological measures" (Brookings, Wilson, & Swain, 1996, p.374), it is evident that more research is still needed on the correlation between cognitive functions and brain oscillations.

The work of Fournier, Wilson, and Swain (1999) emphasized another potential difficulty with the use of EEG in the classification of different mental workload states under multiple task conditions. They showed a statistically significant decrease in alpha 1 (8.2 - 10.1 Hz) for a single-task condition, but they found the measure to be useless under multi-task conditions. The authors assumed that an overall suppression of alpha oscillations when performing multiple tasks produced a floor effect masking workload sensitivity. In addition, an often cited article by Gundel and Wilson (1992) reported about a reduction of parietal and occipital alpha power during a visual Sternberg memory scanning task which was interpreted as being a consequence of the retinal involvement of oculomotor control. Finally, there is evidence from fMRI research indicating a level of cortical activation under multiple task conditions that is less than expected from the conjunction of the activation for each of the two tasks performed alone (Just et al., 2001).

Additional complications for the classification stem from interindividual differences in expertise and task strategy. Fournier and colleagues (Fournier, Wilson, & Swain, 1999) demonstrated that eye blink rate and behavioral measures were sensitive to training. Fairclough, Venables, and Tattersall (2005) found that the level of expertise had a profound influence on the psychophysiological response of an operator to high task demand as demonstrated for respiration rate and supression of alpha activity in a multitask setting (Multi-attribute Task Battery).

It has been demonstrated that the electrophysiological response may vary considerably across subjects. Wilson and Fisher (1995) used topographical information of the EEG to classify fourteen different cognitive tasks and they showed that the use of individual subject EEG patterns had a great advantage over the use of group derived bands. Especially the EEG alpha band has been shown to hold an individually unique signature that may vary with age, memory performance, and attentional demands (Klimesch, Schimke, Ladurner, & Pfurtscheller, 1990; Klimesch, Schimke, & Pfurtscheller, 1993). Therefore, an individual alpha frequency (IAF) adjustment has been suggested for the EEG data analysis. Klimesch and colleagues (e.g. Klimesch, 1996) introduced their method of individual alpha frequency band adjustment which found a wide recognition throughout the literature. Following this procedure, they analyzed event-related band power and showed that the upper alpha band (IAF to IAF + 2 Hz) reflected the processing of semantic information while lower-1 alpha (IAF - 4 Hz to IAF - 2 Hz) was involved in attentional processes. The lower-2 alpha band (IAF - 2 Hz to IAF) was demonstrated to reflect expectancy (Klimesch, Doppelmayer, Röhm, Pöllhuber, Stadler, 2000; for review see Klimesch, 1999). Based on these findings, the EEG frequency spectra in the present work have been individually adjusted (see section 3.3.3.1 for details). Especially in an open environment like real traffic driving, the consideration of individual differences, besides a well-prepared experimental protocol and careful artifact rejection, seems inevitable to obtain a good signal to noise ratio.

2.5.3.2 Paradoxical findings for the alpha band in basic laboratory research.

Basic laboratory research aimed at a deeper understanding of the cognitive functionality of the EEG alpha rhythm. Although great theoretical advances have been made, the topic is still debated today.

Berger (1929) observed rhythmic brain waves around 10 Hz that he termed alpha waves and that were maximal when the person sat relaxed and with eyes closed. Following this observation, it has been concluded that alpha oscillations reflect a relaxed and unoccupied brain. Additional support has been provided by studies that report about a task-related decrease in alpha power during visual stimulation and scanning tasks over occipital electrode sites (e.g. Mann, Sterman, & Kaiser, 1996). Gevins and colleagues (Gevins et al., 1997) found a relationship between increasing memory load in an n-back task and an increase in frontal midline theta activity as well as a decrease in posterior alpha activity (see Gevins et al. 2003 for review). Similar effects are known over sensorimotor areas during movement or somatosensory tasks (see Pfurtscheller, Stancak, & Neuper, 1996 for a review on idling). However, the simple statement that the alpha band activity may represent cortical idling has been widely discussed and criticized (Palva & Palva, 2007). In her review, Rippon (2006) states: "This term [cortical idling] is perhaps unfortunate since it implies no activity, making reports of alpha ERS [event-related synchronization] in successful or skilled task performance seem paradoxical. However, an alternative interpretation of ERS as a measure of the interruption or inhibition of thalamocortical information transfer and thus the reduction in activity of task-irrelevant networks may explain this apparent paradox." (p.253).

In contrast to the idea of "cortical idling", evidence has been provided for a general framework where alpha power amplitude mirrors a level of cortical inhibition (Ray & Cole, 1985a, 1985b; Klimesch, 1996; Klimesch et al., 2007; Pfurtscheller, 2001, 2003). Ray and Cole (1985b) investigated tasks that involved internally directed attention, such as mental arithmetic, word production, mental rotation, and visual imagery and they observed an increase in alpha power as a consequence of rejection of sensory information intake. They explain their results within the framework of an earlier proposed intake-rejection hypothesis (Lacey, 1967). This hypothesis is based on the differentiation between sensory "intake" tasks that involve externally directed attention versus non-sensory "rejection" processes that require internally directed attention. Examples for internally directed tasks are mental arithmetic, mental imagery, and working memory tasks. The core idea of the hypothesis is that in order to facilitate tasks that involve internally directed attention, incoming sensory information needs to be rejected. Empirical evidence in support of this view has been reported by Klinger and associates (Klinger, Gregoire, & Barta, 1973) who saw that imagination elicited increased alpha oscillations and Schupp and his colleagues (Schupp, Lutzenberger, Birbaumer, Miltner, & Braun, 1994) who observed lower alpha band power for perceptual tasks than for mental imagery.

The idea that sensory information intake is rejected was advanced by Klimesch (1997, Klimesch, Doppelmayr, Schwaiger, Auinger, & Winkler, 1999; Klimesch et al., 2000) and Pfurtscheller (2001, 2003) who incorporated it into a wider framework. According to their theory, regions of active neural processing are marked by a suppression of alpha band power, i.e. event-related desynchronization (ERD) while task-irrelevant regions are actively inhibited which is reflected by increased alpha, i.e. event-related synchronization (ERS). Pfurtscheller (2003) named this phenomenon the focal ERD/surround ERS. Cooper et al. (Cooper, Croft et al., 2003) simply described the same mechanism as the surrounding "doughnut of alpha synchronization or inhibition" (p.66). They also suggested that the greater the task demands, the more inhibition is required, and the greater the synchronization that can be observed.

A series of well designed experiments were employed and strong empirical evidence has been presented that speak for the role of alpha oscillation in the active inhibition of task-irrelevant cortical areas. Jensen and co-workers (Jensen et al., 2002) tested the electrophysiological response during the 2.8 s retention interval in a Sternberg working memory task. They varied task load by using different memory set sizes (2, 4, and 6). The results showed a prominent alpha peak at 9 - 12 Hz over posterior and bilateral central regions for 2.5 s of the investigated interval. As demonstrated with 10 subjects, alpha power increased with increasing task difficulty. The authors assumed that neural generators in the parietal-occipital fissure, i.e. the classical source of oscillations in the alpha band were responsible for the parietal enhancement and that sources in the somato-motor cortex could have been responsible for the lateral effect. Jensen et al. considered previous work by Gevins and colleagues (Gevins, Smith, McEvoy, & Yu, 1997) who showed contrary results: a clear decrease in alpha power with enhanced task demands in a n-back working memory task. In response to these results, Jensen et al. argued that subjects performing the n–back task were most likely employing a visual strategy to solve the task. Therefore, increased task demands to the visual system with memory load in the n-back task would explain the posterior decrease in alpha activity. Jensen and his colleagues assumed that the memory-dependent alpha elicited in their own experiment should most likely be generated by the same sources like the sources at parietal-occipital and somato-sensory motor areas that generate the 9 - 12 Hz rhythm at rest (Hari, Salmelin, Mäkelä, Salenius, & Helle, 1997). With respect to previous results from a visio-spatial attention task (Worden, Foxe, Wang, & Simpson, 2000) they speculated that the increase in alpha activity over somato-motor areas would show the disengagement of the motor system during the retention phase. They illustrate their idea by referring to the somehow plausible example of a person freezing in their movements when performing highly concentrated thinking. Nevertheless, Jensen and colleagues left the question open to further research whether enhanced alpha band activity represents the consequence of active inhibition or whether these oscillations have a direct function in memory maintenance. In the light of these and subsequent similar results (Busch & Herrmann, 2003, Sauseng et al., 2005), some authors have preferred the view that alpha oscillations are essential in maintaining neural representations of memorized items (Halgren, Boujon, Clarke, Wang, & Chauvel, 2002; Palva, Palva, & Kaila 2005).

Sauseng et al. (2005) advanced existing concepts by investigating changes in the EEG alpha band during top-down processing in a visuospatial working memory task. They compared a working memory retention condition with a condition in which the retained items had to be additionally manipulated and thus reflected enhanced top-down processing. Their results showed that in the additional top-down manipulation condition alpha power decreased at occipital and increased over prefrontal electrode sites. The authors also observed an equilibrium between prefrontal and occipital regions with respect to absolute alpha power as well as a stronger functional coupling between these areas. Moreover, alpha latency shifts could be demonstrated from prefrontal cortex to primary visual areas that could possibly reflect a controlling influence. Sauseng et al. pointed out that their data neither support EEG alpha as a mere idling rhythm, nor do they support the idea that alpha synchronization would reflect general or global inhibition of task-irrelevant neural circuits when working memory load is increased. They argued that their results would be in line with the idea of selective top-down inhibition needed for concentrating on an ongoing task and discarding any distraction from potentially new activities. The authors suggested that under alpha synchronization cortical areas are in a less excitable state, but that a few selective processes would survive inhibition. In the case of prefrontal alpha activity, this would mean that these areas operate (top-down) to control other areas while simultaneously blocking potentially disturbing external input.

Cooper, Croft et al. (2003) reported data of two experiments showing that alpha amplitudes were greater during both increased task load and during internally directed as opposed to externally directed attention. The authors operationalized internally and externally directed attention by using sensory-intake vs. mental imagery paradigms. Externally directed tasks involved the presentation of a sequence of stimuli from one of three different modalities (acoustic, haptic, and visual). To manipulate task difficulty, subjects had to either simply attend to the stimuli or answer two questions, one being easy and one being more difficult to answer (only one type of question was used in the second experiment). The internal tasks consisted of the subjects imagining the sequences of stimuli under identical task difficulty manipulations. Based on their results, Cooper and colleagues also argued against the notion that alpha would be a marker of cortical idling. The increase of alpha amplitude during imagination as opposed to sensory-intake tasks was particularly evident over electrode sites near visual cortical areas, e.g. Oz and O2. These sites have also been known as reflecting activity in the parietal-occipital and calcarine sulci where the location of alpha generators has been assumed (Hari et al., 1997). Moreover, since the effect of task load did not vary with the direction of attention (Experiment 1) their data support the idea that alpha synchronization reflects inhibition of non-task relevant cortical areas irrespective of the direction of attention. Converging evidence comes from similar studies that investigated the performance of internal tasks like mental arithmetic (Palva, Palva, & Kaila 2005) and mental imagery (Hari et al., 1997) and that successfully demonstrated an increase in alpha power amplitude.

In a later follow-up study, Cooper et al. (2006) addressed the argument that the direction of the alpha band power effect may simply be a matter of methodology, i.e. a result of whether induced (event-related) or evoked (phase locked) alpha band power has been calculated (Klimesch et al., 2000). Klimesch et al. proposed that task-related increases in alpha power are due to evoked activity alone. Based on their work in which they calculated both measures as proposed by Kalcher and Pfurtscheller (1995), Cooper and his colleagues had to reject this objection. By using the identical experimental paradigm as before (Cooper, Croft et al., 2003) they replicated their earlier results, i.e. they found increased alpha power during internally as opposed to externally directed attention. Their results have shown that these task-related increases only derived from induced activity. No experimental effects for evoked activity could be observed in any of the examined alpha sub-bands. The authors offered a possible explanation based on the work of von Stein and Sarnthein (2000) who introduced the idea that alpha activity may be regarded as a an indicator of top-down processing 'par excellence'. Following this, the greater alpha activity involved in internally directed attention tasks may be attributed to greater top-down requirements as opposed to the requirements related to sensory intake tasks.

In conclusion, although some seemingly paradoxical findings still remain a matter of debate (Cooper et al., 2006) evidence has been provided that an increase in alpha band power indexes cognitive processes that relate to working memory and selective attention. Alpha synchronization is regarded to reflect inhibitory processes serving to increase signal to noise ratios and to block out incoming external distractors. In this line stands evidence reporting that internal tasks produced higher alpha power than externally directed tasks due to higher inhibition efforts of distractors in order to focus on the task. Moreover, increased alpha synchronization is observed with increased task demands. In their review on EEG alpha oscillations, Klimesch, Sauseng, and Hanslmayr (2007) proposed that the attentional spotlight may be physiologically implemented through alpha power changes. Task-relevant areas are released from inhibition through event-related desynchronization while task irrelevant regions are suppressed by large alpha oscillations. In a short review of the ongoing debate, Palva & Palva (2007) suggested that the alpha band would be an integrative part of a wide network of alpha, beta, and gamma frequency band oscillations that are involved in unified cognitive operations. They identified alpha amplitude and alpha phase dynamics to play a direct role in mechanisms of attention and consciousness (also see Ward, 2003).

The most obvious implications for the present work can be derived from the findings of the research team around Cooper (Cooper, Croft et al., 2003; Cooper et al., 2006) who investigated internalized attention. They found an increase in alpha power related to the active inhibition of irrelevant brain areas. As reviewed earlier (see section 2.4.2), human performance studies have shown that drivers' mental workload based on in-vehicle distraction can be described in terms of a central processing bottleneck. Recarte and Nunes (2003) named the phenomenon endogenous distraction. Drivers engaged in a conversation are in a dual-task situation in which time sharing can be extremely difficult, i.e. they need to prioritize one task over the other. Driving is a continuous task that requires constant processing of incoming visual-motor information. Based on the above presented laboratory results, it is hypothesized that to increase signal to-noise ratio for the secondary task, input from the driving task is actively inhibited. Therefore, an increase in EEG alpha power should be observed under high mental workload induced by a secondary task during driving.

2.5.4 ERP Indices of Mental Workload

ERPs have been used in numerous research studies and there is a comprehensive knowledge about how they relate to mental workload and attention. For this reason, this type of measure was used as additional data source in the experiments of this thesis. Although ERPs have certain disadvantages (e.g. the need for an average over multiple signals), they can be used as a reliable control measure, especially when the reliability of other measures may not yet have been satisfactorily established. In this section only the P300 and the N100 will be presented since they were in the focus of the data evaluation in this thesis' experiments. For a description of the slow tonic negativity that has also been related to mental workload please refer to overview articles written by Manzey (1998) or Kok (2001).

2.5.4.1 Mental workload reflected by the amplitude and latency of the P300.

The P300 is a positive deflection of the EEG that appears approximately 300 - 400 ms after stimulus presentation. The component's latency has been reported to vary between 250 ms to 900 ms and to show an amplitude of 5 - 40 µV for auditory and visual evoked potentials (Patel & Azzam, 2005). After its introduction by Sutton and colleagues (Sutton, Braren, Zublin, & John, 1965), the P300 has become the most frequently used ERP component, which may also be due to its relatively large amplitude that allows an easy measurement (cf. Coles, Smid, Scheffers, & Otten, 1997, p.95).

Two subcomponents have been identified: the "classical" P3b over parietal electrodes and the P3a which can be recorded over frontal electrode sites and which is characterized by shorter latencies. In the following the term P300 refers to the "classical" P3b component if not stated otherwise. The general interpretation is that the P300 might reflect broad recognition and memory-updating processes. In this way, the P3b seems to relate to match/mismatch with a consciously maintained memory trace and the P3a represents a passive comparator (Näätänen, 1990). It has been proposed that regardless of attentional status (target or nontarget) the P3a is elicited by the more infrequent stimulus (Katayama & Polich, 1998).

In addition, a third subcomponent, the "Novelty P3" has been debated. The Novelty P3 was demonstrated in a modified oddball task in which it was elicited by rare and completely unexpected stimuli (Courchesne, Hillyard, & Galambos, 1975). Subsequent research identified the Supplementary Motor Cortex (SMC) or cingulate gyrus as the neural generators of this component (Dien, Spencer, & Donchin, 2003). However, Simons and colleagues (Simons, Graham, Miles, & Chen, 2001) used a factor analysis of the ERP waveforms to proof that the P3a and the Novelty P3 are identical (for a review on P3a and P3b refer to Polich, & Criado, 2006).

It has been shown that the amplitude of the P3b component can be influenced by different stimulus characteristics such as presentation probability (Duncan-Johnson & Donchin, 1977; Lorist, Snel, Kok, & Mulder, 1994), stimulus sequence (Duncan-Johnson & Donchin, 1983), stimulus quality, attention, and task relevance of the stimulus (Coles et al., 1997) as well as emotional context (Naumann, Maier, Diedrich, Becker, & Bartussek, 1997). Neural generators of the P3b have been well studied. It has been assumed that it is generated in the hippocampus and temporal lobe. This view found support from lesion studies (Yamaguchi & Knight, 1991) demonstrating that after damage to the temporal-parietal junction the P3b was no longer elicited. Similar results were obtained from invasive cerebral electrode recordings (Smith et al. 1990).

Two different theories were proposed to explain the cognitive processes leading to the elicitation of a P3b. Donchin and Coles (1988) provided evidence for the idea that the P3b reflects context updating, i.e. it reflects the evaluation of a stimulus with regard to an internal environmental model (see also McCarthy & Donchin, 1981). On the other hand, it has been suggested that the P300 might rather be involved in activity of memory trace remodeling after the detection of a target stimulus ("context closure theory", e.g. Verleger, 1988).

For the visual (Herbert, Gordon, & McCulloch, 1998) and for the auditory (Zenker & Barajas, 1999) modality it could be demonstrated that the P300 is observable under active and passive oddball conditions, i.e. it became apparent even if subjects did not actively respond to the rare stimulus events. However, the P300 amplitude effect might be reduced by habituation. Ravden and Polich (1998) observed that in an active visual oddball task the P300 amplitude was decreased if target and distractor stimuli were presented equally often.

The P300 has been widely used to study the effects of mental workload on cognitive function. As described in more detail in section 2.3.3 above, secondary loading task paradigms are based on models that assume a shared limited resource pool. By the use of an additional loading task, resource demands of a primary task become apparent. It has been assumed that the amount of resources for the secondary task varies inversely proportional to the resource demands of the primary task and that this reflects the level of mental workload induced by the primary task. In line with resource theory, various experiments have shown that the P300 amplitude can be manipulated by task demand conditions (for review see Manzey, 1998).

Wickens, Kramer, Vanasse, and Donchin (1983) showed the reciprocity of the P300 amplitude for concurrent primary (tracking) and secondary task events (auditory probe, visual probe, or visual probe embedded in tracking task). They showed a reduction in P300 amplitude under dual task conditions compared to a single task control condition. Moreover, it has been shown that for the secondary task embedded into the primary task, the amplitude of the P300 varied directly proportional to the task demands of the primary task. If the secondary task was not part of the primary task, the P300 amplitude decreased with increasing task demands from the primary task.

While numerous research studies could demonstrate the sensitivity of the P300 amplitude to perceptive-cognitive demands, results have been published (e.g. Isreal, Chesney, Wickens, & Donchin, 1980) showing that this ERP component did not reflect the changing motoric demands of a task. Although other more recent studies (e.g. Kamijo, Nishihira, Higashiura, & Kuroiwa, 2007) indicated that the P300 could be manipulated by acute exercise intensity, this important observation prepared the ground for subsequent research in complex close to real-world studies involving inevitable motor influences.

Wester and her colleagues (Wester, Böcker, Volkerts, Verster, & Kenemans, 2008) tested subjects in a driving simulation while they performed a secondary auditory oddball task including standard and deviant tones as well as novel sounds. Their results showed decreased P3a amplitudes for the passive oddball task condition when it was performed simultaneously to the lane-keeping task. For the same condition performance measures were completely unaffected for both tasks. With their data, the authors could demonstrate a high sensitivity of ERPs to additional task demands in a dual task paradigm.

Strayer and Drews (2007) analyzed ERPs to test their inattention-blindness hypothesis in a driving simulation (see also section 2.4.2.1 above). They evaluated the P300 time-locked to the onset of a pace-car brake lights in both single- and dual-task conditions and they demonstrated a 50% reduction in amplitude when talking on a cell phone (dual-task condition) compared to when not talking on a cell phone (single task condition). The authors concluded that the source of performance decline under dual-task conditions would be related to a false encoding of the detected objects rather than an impaired retrieval of information from working memory.

In his review, Kok (2001) criticized the conception that the P300 amplitude would provide a straightforward utility for the measure of processing capacity and mental workload. He identified task difficulty and task emphasis (i.e. the subjects' prioritization of one task over the other in a dual task setting) to have opposite effects on P300 amplitude. The author demonstrated in his study overview that in many experimental paradigms these opposite effects could not be disentangled. Moreover, the complex structure of tasks that were used in many dual task studies complicated a direct interpretation of P300 results in terms of related processes and underlying resources.

After all and despite limitations, it can be concluded that it is well established by empirical data from secondary task designs that the P300 amplitude is sensitive to perceptive-cognitive demands in a mentally demanding task (O'Donnell & Eggemeier, 1986; Tsang & Wilson, 1997). Manzey (1998) declared that the P300 has a specific sensitivity and therefore a high diagnosticity for mental workload. As various studies have shown, the P300 is exclusively responding to changes in perceptive and cognitive aspects of mental workload, but it is mostly insensitive to motoric manipulations. This has made it a valuable psychophysiological index for mental workload measurement in complex field settings like driving and air traffic simulations.

2.5.4.2 Earlier components related to attention and mental workload.

Besides the P300 other ERP components may provide further insights on mental workload (Parasuraman, 1990). Since mental workload is a multidimensional measure, it is unlikely that workload is reflected in variations in a single ERP component like the P300. It has been proposed that early components like P1/N1 reflect demands on perceptual resources like those needed in early selection processes, while the P300 is related to perceptual-central resources, i.e. those involved in identifying the semantic category of a stimulus (cf. Kok, 2001). Thus earlier components, notably the N1 and N2, might also be profitably examined (Parasuraman, 1990). Excellent overviews about early ERP components have been written by Luck (2005) and Coles and Rugg (1997). The description here solely focuses on the N1 and N2 since they were most relevant for an interpretation of the results in the empirical part of the thesis. The P1 and P2 have only been rarely used to study mental workload. They seem to be mostly connected to exogenous factors like stimulus contrast (P1) and simple stimulus features indicating infrequent targets (P2).

The N1 elicited by visual stimuli consists of three subcomponents (Luck, 2005) and is usually influenced by spatial attention (reviews by Hillyard, Vogel, & Luck, 1998 and Mangun, 1995). The first subcomponent typically peaks around 100 - 150 ms poststimulus at anterior electrode sites. The two other subcomponents peak around 150 - 200 ms following a stimulus and they are known to be elicited from parietal cortex and lateral occipital cortex. For the latter subcomponent it has been demonstrated that it was larger when subjects performed a discrimination task than when they performed detection tasks. It has been argued that these results might reflect discriminative processing (e.g. Hopf, Vogel, Woodman, Heinze, & Luck, 2002).

Similar distinctions have been made for the auditory evoked N1 subcomponents (reviewed by Näätänen, & Picton, 1987). The first subcomponent peaks over fronto-central sites at a latency of 75 ms. It is generated in the auditory cortex on the dorsal surface of the temporal lobes. The second subcomponent is a vertex-maximum potential of unknown origin that peaks around 150 ms. The third subcomponent is more laterally distributed and also peaks around 150 ms. It is generated by the superior temporal gyrus. Numerous studies have shown that the auditory evoked N1 component is sensitive to attention. Some authors argued that attention effects in the N1 latency range reflected addition of an endogenous component (Woldorff et al., 1993).

With respect to attention there has been an ongoing disagreement in the literature about the functional significance of the N1 amplitude enhancement (Luck, 2005;

Näätänen, 1990). Hillyard and his colleagues (Hillyard, Hink, Schwent, & Picton, 1973) observed a prolonged negative deflection in auditory ERPs to attended stimuli that could be observed in the subtracted difference wave; hence they labeled it negative difference (nd) wave. In their experiments that typically involve a fast presentation rate of stimuli, the effect of attention was apparent in an amplitude modulation of the N1 showing that when sensory input was unattended it was excluded from further processing.

In contrast to this interpretation, Näätänen (1982) regarded the same component enhancement as a consequence of an enlarged endogenous component termed "processing negativity" (PN) elicited by attended-channel stimuli. The PN emphasizes that the component is related to some form of extra processing assigned to attended events on the basis of a preceding selection process. According to Näätänen, the appearance of a N1 amplitude enhancement with attention is the result from a temporal and spatial overlap of a sensory-evoked N1 component and an endogenous PN. In this way, the negative shift that has been added to the unattended ERP is not necessarily time-locked to the N1 component. The author claimed that the PN wave represents activity of neurons specifically engaged in processing the attended stimuli and that these neurons are different from those that generate the sensory-evoked N1 peak.

With regard to the existing research results, Parasuraman (1990) concluded: "As a sign of early selective filtering, the N1 may be particularly appropriate for detecting workload deficits associated with early perceptual processing" (p. 293). The author emphasized two major findings that should be relevant to mental workload situations. First, the N1 is sensitive to attention allocation in dichotic listening tasks, and second, the component is sensitive to divided attention in a multi-channel task. In his experiment, Parasuraman (1978) showed that the N1 amplitude was reduced under divided attention and enhanced in attended as opposed to unattended channels. However, these results could only be obtained under a high presentation rate. A comparison with a slow presentation rate did not reveal any significant results. Based on these and similar results, Parasuarama (1980) claimed that the reduction in N1 peak amplitude under divided attention is due to a prolonged latency of an additional slow negativity which has a smearing effect on N1 peak amplitude.

In another study, Parasuraman (1985) tested subjects in a two channel intermodal divided attention task in which attention allocation between the two ears was systematically varied. Once more, only a fast presentation rate revealed that visual nontarget stimuli elicited an enhanced N160 and P250 with increased attention allocation. A similar enhancement could be observed for the N1 amplitude when auditory stimuli were presented. Parasuraman concluded again that an additional slow negative shift added to the N1 component.

Parasuraman (1990) described a third study from his lab in which a visual task was used to demonstrate that the N160 and P250 were reduced in two divided-attention conditions as opposed to a focused attention condition. Moreover, only for the focused attention condition, he found a decline in the components' amplitudes across time on task. He interpreted this finding as reflecting a vigilance decrement over time. Other studies reported similar N1 variations by using secondary task paradigms to measure mental workload (Kramer, Sirevaag, & Hughes, 1988; Wilson, Fullenkamp, & Davis, 1994).

Another early component, the N2 has also been discussed in the context of mental workload measurement. The following paragraph can only present a brief outline of the findings (for more extensive reviews and discussions please refer to Luck, 2005; Luck & Hillyard, 1994, Mangun & Hillyard, 1997; Patel & Azzam, 2005).

The N2 is typically evoked 185 - 325 ms after presentation of an auditory or visual stimulus that deviates in form or context from a prevailing stimulus. Oddball experiments have shown that the N2 usually appears before the motor response suggesting that it relates to the cognitive processes of stimulus identification and distinction (Hoffmann, 1990).

Similar to the N1, three subcomponents have been identified for the N2. One set reflected involuntary processing, another was evoked through active processing. The N2a (when elicited by auditory stimuli also-called "mismatch negativity" or just "MMN"; Paten & Azzam, 2005) is an anterior cortical distribution that is evoked regardless of whether a deviating stimulus is paid attention to or is being ignored (Näätänen, 1990). The N2b shows a central cortical distribution that can only be observed under conscious attention (Luck, Girelli, McDermott, & Ford, 1997). The N2c is elicited frontally and centrally during classification tasks. In addition, by using lateral and task-relevant stimuli in visual search tasks, it could be demonstrated that the N2pc deflection served as an index of attentional shift in the occipital temporal region of the contralateral cortex.

The MMN represents an automatic novelty-sensing process (Picton, Alain, Otten, Ritter, & Achim, 2000). With a latency of 150 - 200 ms, the auditory MMN is specifically evoked by physically deviant sounds in a repetitive sequence including deviations in pitch, intensity, duration, location, timing, and phoneme characteristics (for review see Näätänen, 1990). In the attended condition, the MMN is usually followed by the P300 (Coles & Rugg, 1997). It has been shown that MMN latencies increased with increasing standard-deviant intensity deflections. This was interpreted as reflecting an elevated cognitive processing requirement for extensive stimulus deviations (Sams, Alho, & Näätänen, 1983). However, there has also been evidence suggesting that under conditions of highly focused attention during dichotic listening tasks the MMN could be modulated by attentional factors (Woldorff, Hackley, & Hillyard, 1991).

It is important to note that the MMN should not be confused with the N2b which is only evoked by task-relevant stimuli and reflects a stimulus categorization process. The N2b amplitude is larger for auditory deviants which is most likely through template mismatch that can be observed for auditory stimuli over central and for visual stimuli over posterior sites (Simson, Vaughan, & Ritter, 1977). However, the question remains open whether auditory and visual N2b components reflect identical neural processing functions (Luck, 2005). Moreover, it has been reported that the N2b related to inferior anterior ERP positivity, i.e. the P2a, which might represent the interaction between areas of salience representation and feature representation (Potts, Liotti, Tucker, & Posner, 1996).

Kramer, Trejo, and Humphrey (1995) aimed at evaluating the sensitivity of ERPs elicited by task-irrelevant probes to variations in processing demands of a real world task like radar-monitoring. They used a simulated radar-monitoring task in which they varied the density and type of targets to be detected while subjects were presented with a series of irrelevant auditory probes that they had to ignore. The authors demonstrated that the N1 and N2 were sensitive to both the introduction of an additional task and an increase in task difficulty. N1s and N2s elicited by deviant tones over electrode Fz systematically decreased in amplitude from the oddball to the low-load radar monitoring to the high-load radar monitoring condition. Similar results were obtained for the N2 component at electrode Cz. Based on the different scalp distributions, the authors concluded that both components, the N1 and the MMN, reflected increasing task demands independently. Moreover, the authors suggested a revision of Näätänen’s (1990) model since their data demonstrated that availability of attention or processing resources seemed to have influenced the effectiveness of the MMN generator. Therefore, they doubted the assumption that the MMN generator would be responsible to warn executive processors about incoming new information independently from current resource demands. Based on their results, the MMN would rather fit to a later stage of processing, i.e. to the stage of early selection processes. They speculated that automatic mechanisms of attention capture might have occurred which could explain why subjects attended to irrelevant probes that they were initially asked to ignore.

Based on the findings that have been reviewed above, it can be concluded that early ERP components are not exclusively sensitive to stimulus-related features, but that they can also be responsive to cognitive processes (e.g. attention). This makes them an important tool for the measurement of mental workload in laboratory and operational settings.

Although the exact neural contributors are still subject to debate, it has been well established that the N1 amplitude is increased with allocation of attention. Findings from dual-task studies indicate that a latency shift in early components like the N1 occurs when multiple tasks are performed simultaneously resulting in attenuation of amplitude. Similarly, N2 subcomponents seem to decrease in amplitude as task demands are enhanced. The MMN component seems to act as an index of automatic feature analysis, i.e. the detection of physical deviance for both attended and unattended stimuli. However as described above, some studies have shown that the MMN amplitude is not insensitive to manipulation of attentional resources. Unlike the MMN, the N2b is not restricted to the auditory modality. It is elicited by a deviation from a mentally stored expectation of the standard stimulus, although it always depends on the events being task-relevant.

Early ERP components like the P1/N1 have been identified to primarily reflect perceptual/sensory demands of early selection processes. Various research works demonstrated that P1/N1 and P3 represent demands on separate pools of “perceptual” versus “perceptual-central” resources respectively (Kok, 1997, 2001). Although both components have been related to mental workload, evidence from focused-attention tasks (Näätänen, 1982) and divided attention paradigms (e.g.

Mangun & Hillyard, 1990) showed that early negativities (N1, Nd) and the P3 operate in a more or less independent fashion.

2.5.5 EEG Artifacts

The EEG measurement technique is known for its high vulnerability to artifacts. In addition, including ERPs in the analysis typically requires the use of specific paradigms. Therefore, an often shared conclusion is that "these two factors make the recording of acceptable EEG and evoked potentials difficult in the real world" (Tsang & Wilson, 1997, p. 431). As a matter of fact, the high amount of EEG artifacts recorded in an open environment like inside a car in real-traffic or on bumpy dirt roads has posed a very special challenge for the analysis of the experiments' data reported in this thesis. Two main reasons have been discussed why EEG artifacts can be especially problematic for any EEG recordings (cf. Luck, 2005, p. 151). First of all, artifacts may greatly reduce the signal-to-noise ratio since they are usually very large and overlay activity related to cognitive processes. Second, if EEG artifacts are more related to one experimental condition than to the other, they may be enhanced due to averaging and thus lead to a wrong data interpretation. For example, it may be that some driving manoeuvers look more demanding than others. However, in this case a straightforward interpretation of data in terms of mental workload may be difficult since these maneuvers also involve an increased amount of muscle activity related to the drivers' movements.

For this thesis, it is inevitable to address these concerns related to EEG field recordings and to discuss the currently available solutions for artifact handling. Although hardly comparable to noise-reduced laboratory conditions, the results of the present work demonstrate that EEG field recordings are nevertheless feasible by using appropriate methods. This section provides an overview of the different kinds of artifacts and about the different rejection and correction possibilities. It helps to understand the special challenges for the analysis of on-road collected EEG data.

The EEG can be contaminated by several types of artifacts including eye blinks, eye movements, muscle activity, skin potentials, and amplifier saturation that have been illustrated in *Figure 11* below. Please note, that alpha waves in the form of spindles as shown in the figure represent a more specific problem related to the investigation of ERPs. Alpha waves usually show a large amplitude that easily overlays the electroencephalographic response to a triggered event and that can seriously distort a signal average. Usually signal epochs showing alpha spindles are excluded from the further analysis. Moreover, an experienced investigator noticing the occurrence of large alpha spindles in a laboratory test would most likely, depending on the experimental paradigm, try to introduce a reactivating break for the subject since this type of signal is known as a clear indicator of fatigue.

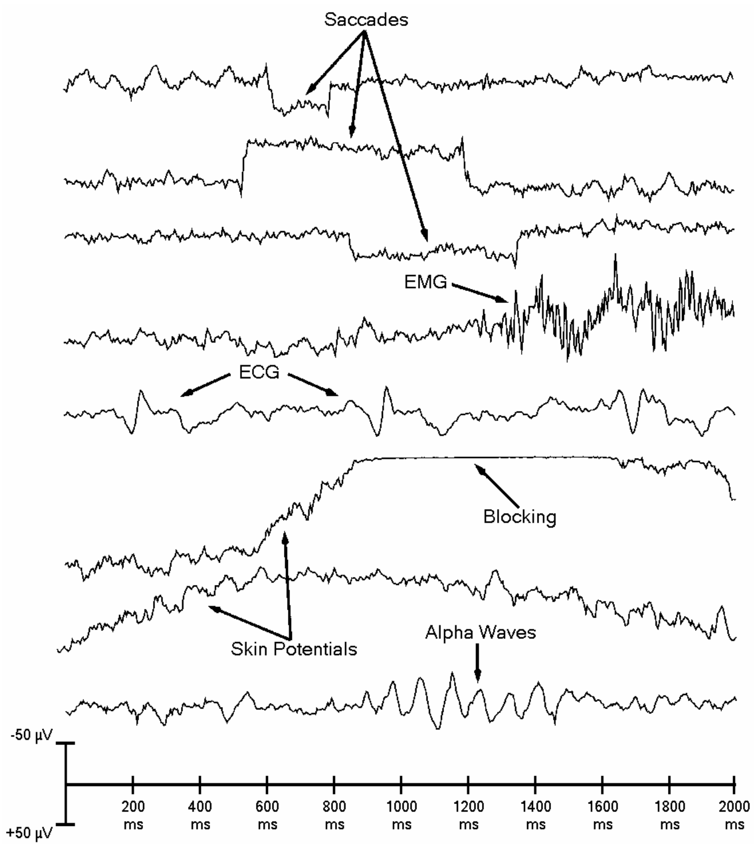

Figure 11. EOG and EEG recordings showing several types of artifacts (adapted from Luck, 2005, p. 164)

2.5.5.1 Eye blinks and eye movement artifacts.

Eye activity is the most common type of artifact and it can be easily observed in an EEG recording. Within each eye there is an electrical voltage gradient stemming from a dipole with its positive end pointing at the front of the eye and its negative pole at the back of the eye (Luck, 2005). The voltage deflections of blinks are caused by movement of the eyelids across the eyes. Eye movements also cause a change of the voltage gradient that can be observed as a positive deflection at sites that the eyes have moved toward. For an efficient detection of eye activity at least one extra electrode is placed directly below the eye. Most electrode setups usually also include electrodes at the outer cantus of each eye and one electrode above each eye (usually Fp1 and Fp2). Many labs include a standard EOG calibration recording at the beginning of their test protocols to account for interindividual differences and to optimize their artifact detection algorithms.

The exemplary eye movements displayed in *Figure 11* above show the characteristic polarity shifts that saccades cause and that can easily be detected in the continuous EEG. The dipoles within the stationary eyes elicit a constant DC voltage that is usually filtered out by the amplifier's high-pass filter. The sloped boxcar-shaped function is frequently observed. It stems from two saccades, one away from a fixation and one returning to it. The impact of eye movements onto the EEG has been intensively studied by Hillyard and Galambos (1970) and Lins and colleagues (Lins, Picton, Berg, & Scherg, 1993a). Their findings revealed a linear relationship between amplitude of the voltage deflection and saccade size by showing that each degree of eye movement can be quantified by the size of the deflection in the EOG.

An example of a true eye blink and a voltage deflection that is not a blink are displayed in *Figure 12*. As can be seen, eye blinks cause a monophasic deflection of 50 - 100 μV and usually show a duration of 200 - 400 ms. An eye blink can be unambiguously identified by the opposite polarity recorded at electrodes above and below the eye.

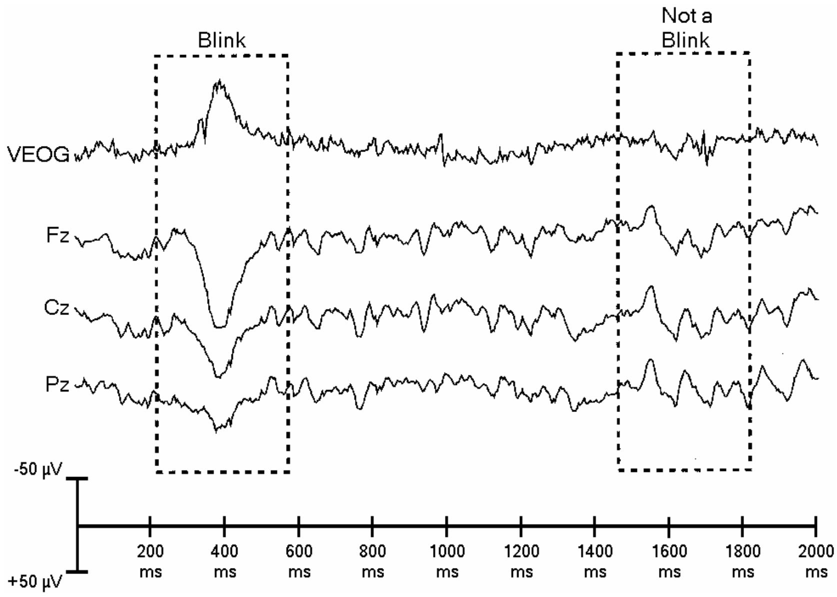

Figure 12. EEG and EOG recordings showing at approximately 400 ms the characteristic signature of an eye blink with opposite polarity below (VEOG) and above the eye. The large voltage deflection at 1500 - 1800 ms doesn't show an inverted polarity over electrodes and has most likely not been elicited by an eye blink (adapted from Luck, 2005, p. 159).

2.5.5.2 Eye artifact rejection and correction methods.

EEG data sections that show eye artifacts usually have to be rejected from the analysis leading to a considerable amount of data loss. Instead of rejecting data afterwards, several ways to already reduce the number of blinks and eye movements during the testing have been proposed. Luck (2005, p. 160 - 161) provided a list of hands-on advices leading from asking subjects not to wear contacts, over to the idea of using shorter trial blocks, and the suggestion that the investigator should provide an appropriate feedback to the participant in order to reduce the amount of eye blinking. Many investigators use a strict test protocol instructing their subjects to blink only during a certain regular trigger event (e.g. Hagemann, 2002). However, it has been argued that this might introduce an additional task for the subjects that could confound the results (Coles & Rugg, 1997, p. 5).

For the offline analysis several ways to automate artifact detection using different types of amplitude threshold procedures have been proposed (for more details see Luck, 2005). Using an automated computer algorithm is probably the most efficient way to handle artifacts related to eye activity. Nevertheless, highly noisy data might limit the application of such algorithms and might make a visual inspection of the EEG inevitable. Especially when using experimental paradigms in which eye activity is integral to the tasks. Moreover, Gratton, Coles, and Donchin (1983) argued that some groups of subjects (e.g. psychiatric patients or children) may not be able to control their blinking and eye movements.

Instead of rejecting and thus losing large portions of data, investigators developed several correction procedures that are based on the estimation and removal of the contribution of blinks and eye movements to the EEG signal (e.g. Gratton, Coles, & Donchin, 1983; for review of six different procedures see Brunia, 1989). Another very successful approach has been developed by Berg and Scherg (1991) who used dipole modeling to extract the ocular activity. However, since the EOG signal always also contains brain activity besides the true ocular activity, subtracting this signal would always lead to a subtraction of cognitive brain activity as well. In conclusion of their review of different correction techniques, Lins and colleagues (Lins, Picton, Berg, & Scherg, 1993b) did not recommend the use of any of the procedures due to the risk of introducing serious distortions into the signal.

Jung and colleagues (Jung, Makeig, & Humphries, 2000; Jung, Makeig, & Westerfield, 2000) successfully used a mathematical procedure called "independent component analysis (ICA)" to remove eye blinks, eye movements, and electrical noise (see also Joyce, Gorodnitsky, & Kutas, 2004). *Figure 13* demonstrates the results of an independent component analysis as it has been performed on a subject's data from Experiment 1 of this thesis.

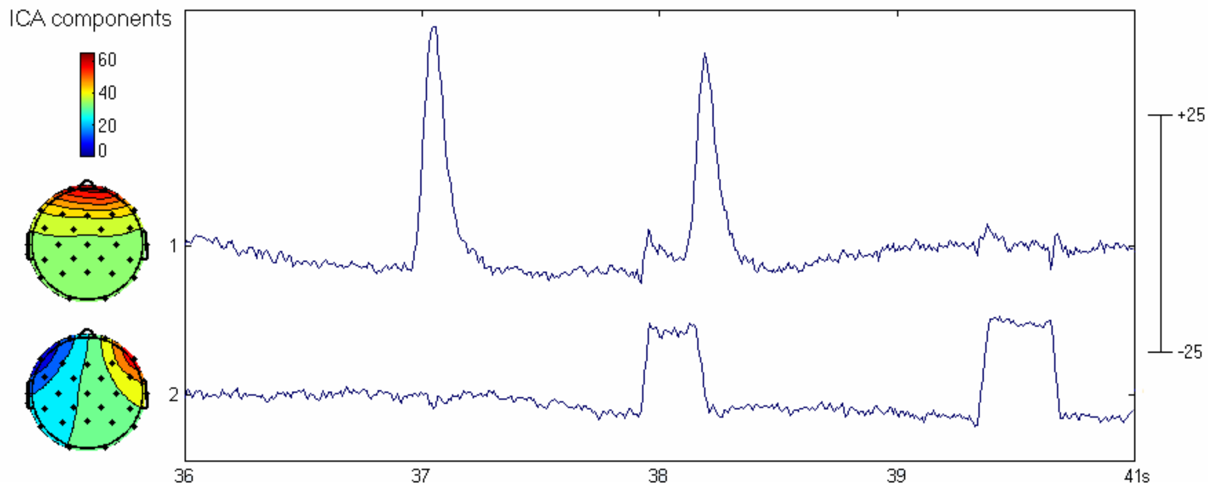

Figure 13. Example of ICA components related to eye activity. The topographic distributions of each component is displayed left. At the right side the components' activations over time are displayed. Please note that this may not be confused with the voltage deflections of the continuous EEG (modified graphics originally produced with EEGLAB, Delorme et al., 2002).

Component 1 shows the typical frontal distribution of an eye blink while Component 2 displays the inverse polarity and the lateral frontal distribution that characterizes eye movement components. Adjacent to the topographic maps, each component's activation over time is displayed which provides complementary information to identify a component. The blink component shows a brief and high amplitude activation while the shifts of activation for component two clearly indicate horizontal eye movements.

Removing eye artifacts with ICA is a promising approach and this method is even more efficient if muscle artifacts and electrode drifts are eliminated beforehand (as done in the analysis of this thesis' data). However, Luck (2005, p. 172) suggested to use some general caution towards this new technique that has not yet been rigorously tested by an independent laboratory. He pointed out three issues that still need to be solved. First, the underlying assumption that artifacts are independent of the time course of ERP activity may not always be true. For example an ERP component and an artifact like a blink may be correlated which would lead to an erroneous artifact correction. Second, for using ICA it has been recommended to use at least seven eye electrodes (Lins et al., 1993b) and to conduct several calibration runs before a test. This imposes some extra effort that has to be considered. Finally, an argument that addresses all correction methods is that sensory input may be caused by blinks and eye movements themselves (e.g. eye movements are accompanied by a sliding of the visual world across the retina which elicits a sensory ERP response). This type of contribution to the signal is not accounted for by correction techniques.

2.5.5.3 Muscle activity.

A classical muscle artifact can be seen in *Figure 11*. Much of the high frequency EMG activity can be minimized by asking the subject to sit relaxed and to avoid unnecessary movements. In laboratory experiments a chinrest is frequently used to achieve a relaxation of the neck muscles. Since theses artifacts are so easily visible in the continuous EEG, a brief demonstration at the beginning of the recording can be useful to increase the subjects' commitment during the experiment. However, even if an experimental protocol is very strict including control for driving behavior (e.g. predefining a speed) and talking (not allowed), an operational environment will never achieve a signal-to-noise ratio that can be compared to a laboratory setting. Sources of artifacts in a driving situation mostly stem from the drivers' muscle activity related to task-specific movements like steering, moving the head to detect passing cars, or even coughing or swallowing. For the two field experiments reported in this thesis the amount of EMG artifacts was substantial. Some authors presented automated detection techniques of EMG with impressive detection rates, e.g. Gevins and Smith (2003) reported for their own data that "algorithms successfully detected 98.3 % of the artifacts, with a false detection rate of 2.9 %, whereas the average expert human judge found 96.5 % of the artifacts, with a 1.7 % false detection rate" (p. 118). However, the distortion in the present data did not allow a satisfactory application of any automated detection. As a consequence, all data were visually screened and EMG activity was manually excluded from the analysis (for more details on the procedure please refer to the methods section of Experiment 1 below). Nevertheless, it has to be admitted that the subjective rejection of EEG data segments involves the risk of adding a subjective bias to the EEG data. Therefore it is important that the investigator analyzing the data is blind about the start and end of experimental conditions in the raw EEG data. In addition, some experience is needed to be able to detect the start and end of muscle related activity or electrode drifts in the continuous EEG raw data.

2.5.5.4 Slow voltage shifts, amplifier saturation, and heart activity.

Slow voltage shifts typically stem from a change in the impedance of the skin or electrodes which leads to changes in the small voltage existing between the deep and superficial layers of skin. Even slight sweating can lead to slow voltage shifts that are called skin potentials in the EEG. Voltage shifts are often the consequence of loose electrodes which can be caused by subjects who lean their head against their headrest or who move around too much.

Slow voltage shifts may sometimes lead to an amplifier saturation or blocking which is characterized by a flat line in the signal (see *Figure 11* above). Setting the gain on the amplifier to a lower level can usually avoid the occurrence of this artifact. If recorded in the EEG, the drifts and amplifier blockings usually need to be rejected. Moreover, if an electrode frequently shows voltage drifts then this electrode may have lost contact to the skin during the recording and the whole electrode needs to be rejected.

Finally, heart activity can be picked up by a misplaced mastoid electrode. If this electrode is used as a reference then the rhythmic ECG signal is observable about once per second in the inverted form in all electrodes. Since this signal is so easily identifiable, ECG activity is usually no significant problem for the subsequent analysis. Generally, the misplaced electrode is quickly corrected during electrode setup.

2.5.5.5 Non-physiological external disturbance sources.

While laboratories should be set up very carefully so that any external electrical source that could possibly distort the EEG signal is minimized (e.g. Bösel, 1996), the car driver already sits inside a Faraday cage. Nevertheless, non-physiological artifacts inside the car could stem from the test equipment like a power line or other sources (e.g. seat heating system). These artifacts are often identifiable by the sudden appearance of waveforms very different from cerebral activity. Line noise oscillations in the EEG recording can usually be eliminated by a notch filter at 50 or 60 Hz and have therefore frequently been used. Other sources of noise inside the car can be very specific to the experimental vehicle and its built-in electronic equipment. Therefore, a test recording and a frequency analysis should always be performed before starting an experiment to check for any distortions (cf. Luck, 2005).

2.6 ECG Indices of Mental Workload

The ECG is recorded from the surface of the skin and it measures differences of an electrical potential elicited by the depolarization and repolarization of the heart muscle. The excitation of the chamber of the heart is represented in a characteristic PQRST signal (see *Figure 14*). Heart rate and heart rate variability have been widely used and established as a standard psychophysiological index for mental workload. Heart rate is usually determined by using the QRS peak and it is measured in beats per minute (Birbaumer & Schmidt, 2006). In the following section some of the main findings are summarized.

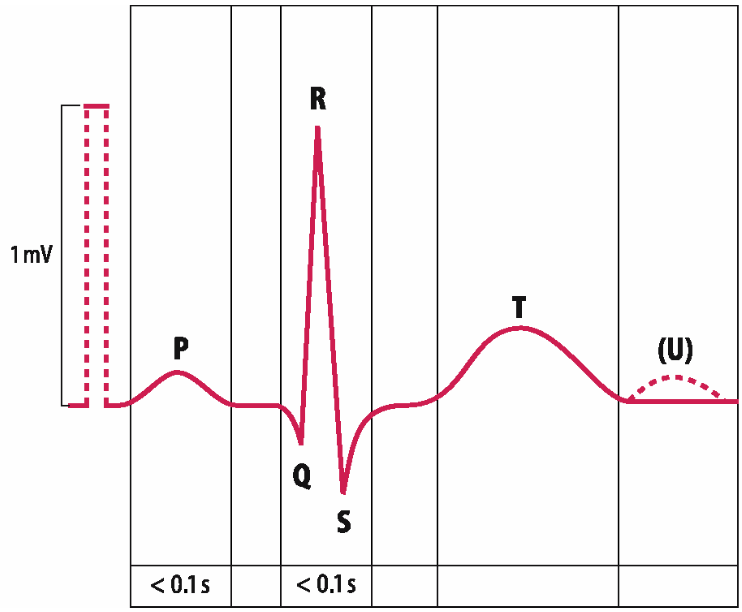

Figure 14. Idealized illustration of a standard signal in a normal ECG. The U wave is not constantly observed. The P wave signalizes the depolarization of the atrial muscles, the QRS complex reflects the depolarization of the ventricles, and the T wave signifies the repolarization of the ventricles. Positive voltage deflections are displayed upwards (modified from Birbaumer & Schmidt, 2006, p. 192).

2.6.1 Heart Rate

The heart rate is easy to obtain and relatively stable against artifacts which made it a popular physiological technique in the assessment of mental workload (Kramer, 1991). As shown by various studies the basic understanding is that the heart rate increases with increasing cognitive and psychomotor demands (for study review see Ribback, 2003). Nevertheless, there have also been a few studies that did not find evidence for such a correlation (e.g. Casali & Wierwille, 1983). As Manzey (1998) pointed out, the findings on the underlying psychophysiological mechanisms have remained unsatisfactorily vague. A relationship has been assumed between an increased cardiovascular activity on the one hand and a heightened cortical energy transformation as well as increased metabolic demands related to high mental workload on the other hand (Backs & Seljos, 1994). But as demonstrated by Carroll and colleagues (Carroll, Turner, & Hellawell, 1986) it is unlikely that increased metabolic demands represent the sole factor adding to an increase in heart rate under cognitively demanding conditions. They tested several cognitive tasks and obtained increased heart rates that could not have been evoked by increased metabolic demands alone.

The heart rate has been regarded as a universal, but relatively unspecific indicator for mental workload. In fact, it has been considered to be sensitive, but not diagnostic of mental workload (Manzey, 1998). Lee and Park (1990) showed that physical demands could have a dominant impact on heart rate measurement. Luczak (1987) suggested that emotional demands stemming from fear of failure or time pressure could be sufficient to evoke an increase in heart rate under high task load. He assumed that the influence of emotional factors and mental workload factors would be indistinguishable. But based on more recent results showing that heart rate varied with mental workload under the exclusion of emotional factors, this idea has been rejected (Ribback, 2003).

Overall, it has to be concluded that heart rate has often been used as an indicator of mental workload since it could provide an easy to obtain and robust measure. Several studies have been reported that used it as a measure for mental workload in car driving (e.g. De Waard, 1996; Hering, 1999), flight simulation (for overview see Roscoe, 1992), and in flight (for overview see Wilson, 1991). Nevertheless, conclusions that are solely based on heart rate analysis in real-world situations should be handled with caution (Tsang & Wilson, 1997) and the use of additional measures is advisable.

2.6.2 Heart Rate Variability

Kalsbeek and Ettema (1963) were among the first to demonstrate that the heart beat becomes more regular under mental workload. They analyzed the variability of the interval lengths between R-peaks in the ECG and demonstrated the sensitivity of heart rate variability, but not the heart rate, to the increased mental demands of a binary choice task. Since then this finding could be replicated with various other tasks (e.g. Backs & Seljos, 1994; Lee & Park, 1990). Different global measures of dispersion have been used and different recommendations with regard to their sensitivity have been given (see Manzey, 1998).

For the present work four standard measures (Baumgartner, 2004) were used: (1) the standard deviation of R-to-R peak intervals, (2) the percentage of intervals with at least 50 ms deviation from the preceding interval, and (3) the mean square root of successive interval differences. An increase in any of these measures indicates parasympathetic activity while the three measures decrease under high mental workload conditions. These measures have been frequently reported throughout the literature. However, it has been argued that their use as a sensitive measure of mental workload might be limited (Manzey, 1998). More sophisticated analysis techniques have been introduced that use spectral analysis of R-R-interval timelines which allowed a more qualified parameterization. Three different frequency components each with a unique functional significance could be identified (Kramer, 1991). The lower frequency band of 0.02 - 0.06 Hz has been related to vasomotor activity involved in the regulation of body temperature. The middle frequency range of 0.07 - 0.14 Hz (also called the "0.10 Hz component" after the main frequency component; de Waard, 1996) has been shown to be involved in the regulation of blood pressure. Although the physiological sources could still not be unambiguously clarified, it has been assumed that parasympathetic and sympathetic influences are displayed within this frequency band. Finally, the oscillations in the high frequency band of 0.15 - 0.50 Hz have been labeled "Respiratory Sinus Arrhythmia" since this measure exclusively reflects parasympathetic influences that are dependent on respiration frequency (Grossman, 1992). Due to this effect, speaking and related changes in the respiration characteristics could lead to serious artifacts in the measurement and should be controlled (Mulder, & Mulder, 1987). But also for the 0.10 Hz component, the influence of respiration could not be excluded (Althaus, Mulder, Mulder, Van Roon, & Minderaa, 1998). Moreover, it has been shown that the temperature of the environment can exert a confounding influence on the outcome of heart rate variability measurements (Razmjou & Kjellberg, 1992).

All three frequency bands have shown a reduction in heart rate variability due to increased mental demands. However, Mulder (1980; Mulder & Mulder, 1981) demonstrated that among the three frequency measures, the middle frequency range of 0.07 - 0.14 Hz was the most sensitive to cognitive task demands. Based on this and subsequent works, Jorna (1992) concluded that the 0.1 Hz component may reliably differentiate between a baseline and a demanding condition, but that it could not be used to differentiate between difficulty levels of a task. Moreover, it has been suggested that the 0.1 Hz component may represent a rather unspecific indicator of mental effort that provides, as the heart rate, no diagnosticity (cf. Mulder & Mulder, 1987). The same authors also pointed out that in order to obtain reliable results for heart rate variability frequency measures, a data size minimum of at least 5 min needs to be available for the analysis.

Due to these limitations and for the sake of simplicity with which the measures could be obtained only heart rate and simple dispersion measures have been evaluated in the experiments of this thesis. They were used as simple additional control measures.

2.7 Summary of the Theoretical Background

The research results described above clearly demonstrate the amount of effort that has already been spent on investigating mental workload during driving. Numerous techniques have been used, but as can be seen from the still ongoing research activities, up-to-date there is no existing technique that would allow a completely reliable and valid measurement of mental workload. In comparison to other methods, psychophysiological measurements demonstrated clear advantages in directness of the measure, unobtrusiveness, and time resolution. A high amount of laboratory studies indicated that the EEG can be used to measure the current cognitive state of a human. It is considered scientifically demonstrated that the EEG alpha frequency band is sensitive to manipulations of mental workload and direction of attention. The extraordinary capability of the EEG to provide a real-time measure gave reason to numerous research efforts that used the EEG in real-world situations. Many of these studies aimed at the development of brain-adaptive man-machine systems. However, the same studies showed the necessity to use multiple data sources. The available spectrum of methods has been expanded in the past decades even allowing the analysis of data from difficult conditions, i.e. highly artifact-prone EEG data from operative environments.

The number of publications that used the EEG to address research questions in driving has been limited. The reasons for this may have certainly been rooted in difficulties due to the complexity of the driving situation as well as the measure’s high sensitivity to artifacts. Nevertheless, especially in the area of automobile development there is an excellent potential for the EEG to use it in the design of systems aiming at the increase of traffic safety. Last but not least, the presented research in the field of cell phone use during car driving demonstrated the heated discussion around the potential risks involved in high mental workload due to secondary tasks in real-traffic driving.[3] In this context, some authors coined the terms "endogenous distraction" and "inattentional blindness" to describe the consequences of natural cognitive processing limits. Up-to-date no work has been published that investigated the impact of cognitively loading secondary tasks onto the driver is mental state by using EEG in real-traffic driving.

In the following three experiments, the drivers' mental workload has been investigated first in the laboratory, then in a field experiment in real traffic, and finally under almost off-road conditions by using combined psychophysiological and behavioral measures. The author explicitly intended to follow an experimental approach that focused on the distraction or mental workload of the driver due to secondary tasks. The complexity of the driving situation has been minimized by using a strict experimental protocol. The aim of the experiments was to mirror the process of attention allocation away from an external driving task, i.e. the situation on the street, towards an internally directed mentally loading task and to verify this process by showing changes in the EEG and in additional other data sources. To underline the generalizability of the observed effect the workload manipulations were conducted over different experimental settings.

[3] Certainly there are not only disadvantages involved in the use of cell phones during driving, e.g. using a cell phone may also help to reactivate the driver under monotonous driving conditions (see chapter 2.1.4 above).

3 Experiment 1

3.1 Aim and Hypotheses

The aim of this experiment was to examine the impact of a mentally straining, conversation-like task during simulated driving onto the alpha power amplitude in the EEG. A laboratory setting with optimally controlled conditions for the EEG recording was chosen for a first examination of the effect. The laboratory experiment aimed at replicating previous results that showed a large amount of alpha power during both increased task load and internally directed attention. However, previous experiments did not investigate a multiple task situation. Specific auditory tasks were chosen that were suitable for usage in a dual-task design, i.e. tasks that could be performed in parallel to a driving simulation. Moreover, the high workload task condition included elements that resembled listening to a conversation partner. Since there are already a high number of investigations examining this topic, the present work can benefit from those results and establish a transfer to the findings in the present experiment. Furthermore, in order to pin down the influence of the simulated driving task onto the EEG and in order to investigate the interaction between secondary task and driving, the same set of tasks were tested in parallel to driving as well as without it. Although the electrophysiological response to mental workload was the main focus of this investigation, several control measures (behavioral and peripheral physiological measures) were included in terms of a multilevel approach (e.g. Manzey, 1998). In this experiment, mental workload was induced by a story listening task combined with a key word detection task. The electrophysiological and peripheral physiological responses as well as the behavioral performances of the high workload phases were compared with a low mental workload condition in which subjects performed a simple tone detection task. The set of hypotheses tested in this experiment is presented in the following.

Hypothesis 1.

Based on previous research results reviewed above, it was hypothesized that subjects performing a mentally straining, conversation-like task try to compensate for the increased demands by reallocating more of their attentional resources to this task. In accordance with EEG experiments investigating internally directed attention an increase in EEG alpha power compared to a less-demanding simple tone detection condition should be observed. The heart rate was expected to be increased and the heart rate variability should be decreased under the more demanding task condition. Longer reaction times and a lower detection rate (d′) were predicted when subjects detect key words as opposed to simple tones.

Hypothesis 2.

The increase in EEG alpha power in the demanding word detection task compared to the simple tone detection task should also be observed when simultaneously performing a driving simulation as primary task. The additional measures (heart rate, heart rate variability, and detection performance) should show the same effects of mental workload as under single task conditions (see hypothesis 1). Moreover, the increase in mental workload should also be reflected by a worse performance in the driving simulation, i.e. an increased deviation from the reference model in the Lane Change Test (LCT) when listening to the stories compared to the detection of tones.

Hypothesis 3.

Regardless of workload manipulation, the performance of an additional driving simulation requires more attentional resources than performing only the auditory detection task alone. This should be reflected in the detection tasks performances, i.e. an overall increase in reaction times and an overall reduction in d′ should be observed. The heart rate should be higher and the heart rate variability should be lower than under single task conditions without additional driving.

With respect to the overall EEG alpha power level under the condition of performing a driving simulation in parallel, an additional increase in alpha power did not seem likely in the light of research showing an overall suppression of alpha oscillations when performing multiple tasks (see section 2.5.3.1). Moreover, the extensive visual stimulation from the LCT compared to the simple fixation during the auditory task should exert its effect that is primarily related to the physical properties of the stimuli. Therefore, a general decrease in alpha power should be observed when comparing the multiple task condition (LCT and auditory task) with the single task condition (auditory task only).

3.2 Methods

3.2.1 Subjects

Twenty-one participants (3 female) volunteered for the study. The subjects' age ranged from 24 to 45 with an average age of 30. After testing all subjects with a German version of the Edinburgh Inventory (adapted from Oldfield, 1971) 1 subject was identified as being left-handed. They were ignorant about the purpose of the experiment and had less or no experience with the EEG procedure. All participants had a driving license and reported normal hearing and visual abilities. After the experiment, each subject received a small gift from the Mercedes-Benz accessory shop (e.g. a coffee mug) as compensation for their efforts.

3.2.2 Experimental Setup

The experiment took place in a 4.7 m x 4.1 m quiet room, equipped with 3 computers. Notebook 1 (Dell, Precision M60, 1.2 GHz) presented the auditory stimuli. Notebook 2 (Compaq Evo N620C, 1.4 GHz) recorded the physiological data (EEG and ECG) as well as responses via a separate number pad. Additional EEG recording equipment was placed on the table to the left of the subjects. The desktop computer (2.3 GHz) was used to present the LCT (Sim v1.2; Mattes, 2003) as well as a fixation cross (small black cross on grey background) during the blocks without driving task. The participants sat at a table with a LCD monitor (19'' diameter, resolution of 1024 x 768 px, approximately 60 cm distance between the subject and the monitor) and keyboard in front of them. With the right hand, subjects used the arrow keys on the keyboard to steer the car in the LCT. With the index finger of the left hand subjects responded to the stimuli of the auditory presented secondary task. The participants listened to the stimuli via earplugs. A microphone was attached to the upper frame of the monitor to record the participant's oral response with a digital voice recorder (Sony ICD-P17).

3.2.3 Physiological Data Acquisition

The participants' EEG was recorded with an electrode cap (EASY CAP), containing 29 Ag/AgCl sintered electrodes arranged according to the standard 10 - 20 system (Jasper, 1958). One ocular electrode was placed about 2 cm below the right eye. Two electrode positions in the cap had been modified to serve as horizontal eye electrodes, i.e. electrodes F7 and F8 were positioned approximately 0.5 cm closer to the eyes. Two ECG electrodes were used which were placed according to a modified version of the standard extremity recording setup. One electrode was placed at the upper end of the sternum about 2 cm to the right and towards the heart. The second electrode was located to the right of the acromastium, approximately below the 4th rip. This placement allowed the recording of the signal next to the apex of the heart.

The EEG was recorded relative to a reference electrode placed at FCz. The ground electrode was integrated into the electrode cap and placed at position AFz. Data were digitized at 500 Hz with a resolution of 0.1 (range = +/- 3.2768 mV). A time constant of 10 s (low cutoff = 0.016), a high cutoff filter at 1000 Hz (raw data saving filter = 100 Hz) and a notch filter at 50 Hz have been applied. Impedances were kept below 5 kOhm. All data were recorded using Brainproducts recording hardware and the Brain Vision Recorder 1.03.0002 software. The data were processed using the EEGLAB (v. 4.512, Delorme & Makeig, 2004) toolbox under MATLAB (v. 7.2.0.232, R2006a).

3.2.4 Task Description and Stimulus Material

Subjects performed two different types of tasks in this experiment: a LCT and a workload-inducing auditory detection task with two levels of difficulty. Every participant was instructed not to talk during the experiment and to avoid unnecessary movements not related to the tasks in order to avoid artifacts in the EEG.

Figure 15. Screenshot of the Lane Change Test (LCT) designed by Mattes (2003). Double arrows on the signs instruct the driver to change the lane.

The LCT is a widely used tool that has been originally developed for driving research to evaluate driver distraction (Mattes, 2003). The task consists of a simple driving simulation in which traffic signs along the drive require the subject to perform lane changes. In this experiment, a view directly out of the front windshield was displayed (see *Figure 15* above). Standard driving dynamics were used that simulated the driving of a Mercedes-Benz SLK roadster car model. Subjects drove in continuous tracks at a constant maximum speed of 50 km/h on a three-lane road. A maximum total of 10 tracks with a total length of 29383 m and a mean length of 3265 m (range = 3175 m - 3335 m) were available in the program. At the end of a track, subjects drove a curve and automatically entered the following track. 150 m after the starting point of each track the participants saw the first sign. Then the next signs came into view as the participants continued driving. Each sign showed three symbols which represented the three lanes of the road, i.e. two x's and one double-arrow. Subjects were supposed to change the car to the lane indicated by the double arrow as quickly and as accurately as possible. In each track a total of 18 signs were presented. The mean distance from sign to sign was 150 m (range = 140 m – 188 m) so that the mean appearance rate was 10.78 s (range = 10.08 s – 13.54 s). Signs were readable 40 m, i.e. about 2.9 s before passing them. Each of six possible lane changes occurred exactly three times on each track and the order of signs was randomized in each track to avoid learning effects. In one half of the experiment, subjects performed the LCT in parallel to the auditory task. During the other half, subjects looked at a fixation cross presented on the screen. The order of these two sessions was counterbalanced across subjects.

The auditory workload task included two conditions: an easy tone detection task and a difficult story listening task which was combined with a word detection task. In the simple tone detection task a high (500 Hz) and a low tone (400 Hz) were presented to the subjects. They used their left index finger to respond to the high tones by pressing the "0" key on a separate number pad in front of them. Participants were instructed to press the button as quickly and as accurately as possible. The tones' durations were 70 ms with a rise and fall of 10 ms. The stimulus material had been recorded beforehand and was presented in a pseudo-randomized order with an inter-stimulus interval (ISI) of 4 s. "Brown" noise was generated and mixed into the second sound track to obtain a constant background noise level equivalent to the general noise level of the story blocks that were used in the word detection task. At the beginning of the experiment the sound volume of the stimulus presentation was individually adjusted until the participant confirmed that the stimuli were well audible. Eight recordings of 3-min length were generated and in each presentation the first tone was presented after an initial pause of 1 s. Due to the offline generation of the stimulus material the frequency of high relative to low tones in each block was only approximately 50 : 50 (low tones: 47 % - 64 %, high tones: 36 % - 53 %). The presentation of the eight recorded blocks was randomized over all subjects. Blocks of the signal-response task always followed a 3-min story block.

For the word detection task 3-min sequences of the commercially available German audio book “Seven Years in Tibet” (“7 Jahre in Tibet”, Harrer, 1952) were extracted and for each sequence a keyword was selected. All sequences were narratives read by one male reader, i.e. no other voices and no sound effects were audible. Selected keywords were frequent (17 - 22 times) in the text and were mostly articles like the German word “die” (engl. “the”) and personal pronouns like the German word “wir” (engl. “we”). The participants had to press the “0” button on the separate keypad whenever they detected a keyword in the ongoing story. In addition, subjects were instructed to carefully monitor the content of the story. After each block, subjects were asked a multiple choice question (three alternatives) about the content of the story that they just heard. The answers were recorded with a microphone for later offline analysis. As stated above, for both the tone and the word response tasks, subjects used their left index finger to press the “0” key on an additional number block that wrote the responses directly into the marker channel of the EEG recording. The alternating sequence of story block and signal-response block presentation was randomized over subjects. Four out of eight blocks of each block type were presented simultaneously to the LCT.

3.2.5 Experimental Procedure

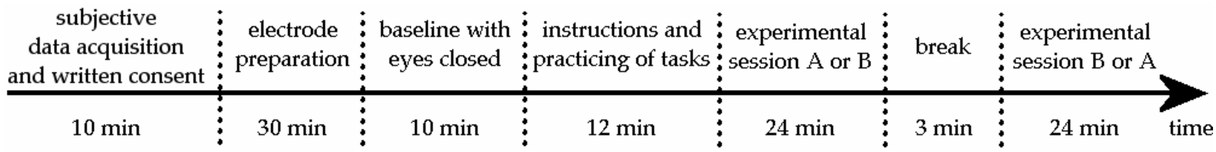

Figure 16. Schematic overview of the experimental procedure. Experimental session A and B either refer to a session with simultaneous Lane Change Test or a session in which participants looked at a fixation cross. The order of sessions was counterbalanced across subjects. In each session four 3-min blocks of listening to a story and four 3-min blocks of listening to tones were alternately presented. See text for details.

Figure 16 provides an overview of the experimental procedure. Before starting with the experimental tasks, subjects were asked to sign an informed consent and to fill out a questionnaire (see section 3.2.1). Next, the physiological recording was prepared which included a brief artifact demonstration (eye blinks, eye movements, and jaw clenching) to increase compliance with the rigid experimental protocol (no talking, avoid unnecessary movements etc.). Every experiment started with a 10-min baseline recording during which subjects had to sit alone in the room, close their eyes, and try to relax. After the baseline recording, the tasks were explained to the subjects and the high and low tones of the detection task were presented until subjects could correctly identify them. Next, participants had 12 min to practice the LCT together with the auditory tasks (two story blocks alternating with two blocks with tones) or until they confirmed that they understood the task and were comfortable with the controls. After this test run, all subjects reported to be comfortable with the tasks and that they were able to continue. Recording sessions were either conducted between 9 and 11 am or 1 and 4 pm and the participants were randomly assigned to either the morning or the afternoon session. The experiment consisted of two experimental sessions (driving simulation vs. fixation). The order of these sessions was counterbalanced across subjects. In each session four three-min story blocks and four three-min tone detection blocks were alternately presented. The sequence of tasks was held constant between blocks, i.e. each block started with the word detection task. However, eight different sets of stimuli were employed for each type of task and their sequence of presentation was pseudo-randomized across subjects.

3.2.6 Experimental design

Table 3. Overview of the experimental design of Experiment 1.

	Mental Workload Manipulation	
	Story listening task with word detection	Simple tone detection task
No driving simulation	Session A	
Simultaneous Lane Change Test	Session B	

Note. All 21 subjects were tested in each of two experimental sessions in a complete repeated-measures design. Each session consisted of four 3-min blocks of each workload task.

Every subject was tested in each of two experimental sessions, i.e. either with or without performing the LCT simultaneously (see *Table 3*). Two main factors with two levels were tested. Factor 1 consisted of the mental workload manipulation by the type of auditory task (difficult story listening and word detection task vs. easy tone detection task). The second factor was labeled "number of tasks" and it describes conditions in which subjects performed the auditory tasks alone or in parallel to the driving simulation. Although subjects were instructed to detect words and carefully monitor the content of the story, the high workload task condition is simply referred to as "word detection task" in the following text for reasons of easier readability. The various behavioral, neurophysiological, and peripheral physiological measures that represented the dependent variables in Experiment 1 are listed in *Table 4*.

Table 4. Overview of the different dependent variables of Experiment 1.

Behavioral performance	
	Lane change performance
	Mean deviation from reference model
	Standard deviation of steering angle
	Percentage of correct answers about story content
	Reaction times for word or tone detection
	Accuracy (d’) for word or tone detection
EEG	
	Individual and task-specific alpha power amplitude
ECG	
	Heart rate (bpm)
	Heart rate variability
	Standard deviation of N-N intervalls (SDNN)
	Percentage of N-N intervals with at least 50 ms deviation from the preceding N-N interval
	Root mean square of successive N-N interval differences (RMSSD)

3.2.7 Data Analysis

The following section describes the details of the behavioral, neurophysiological, and peripheral physiological data analysis. The section concludes with some considerations on the post-hoc calculation of statistical power for the three experiments of this thesis.

3.2.7.1 Behavioral data analysis.

Lane change evaluation.

Sections 100 m after the last sign and 150 m at the beginning of each track included curves which were excluded from the analysis. Subsequently, the log-files were analyzed using the LCT Analysis (v1.99) tool which is part of the LCT software package. To evaluate the driver's driving performance, the deviation between a normative model and the actual course of the subject along the track was calculated. Reference lanes were individually generated and fitted to each subject. The program's default interval of 100 m before and 20 m behind each sign was used for the calculation. In addition, the standard deviation of the steering angle was also computed. A symbolic example of the normative model and the driving data is shown in Figure *17*. An increase in deviation is used as an indicator of an increased level of driver distraction (Mattes, 2003). The mean values for each experimental block were exported to SPSS (v. 12) for statistical analysis of workload effects. 1 subject out of 20 was excluded from the analysis due to missing data for half of the blocks. Another subject was excluded due to implausible values.

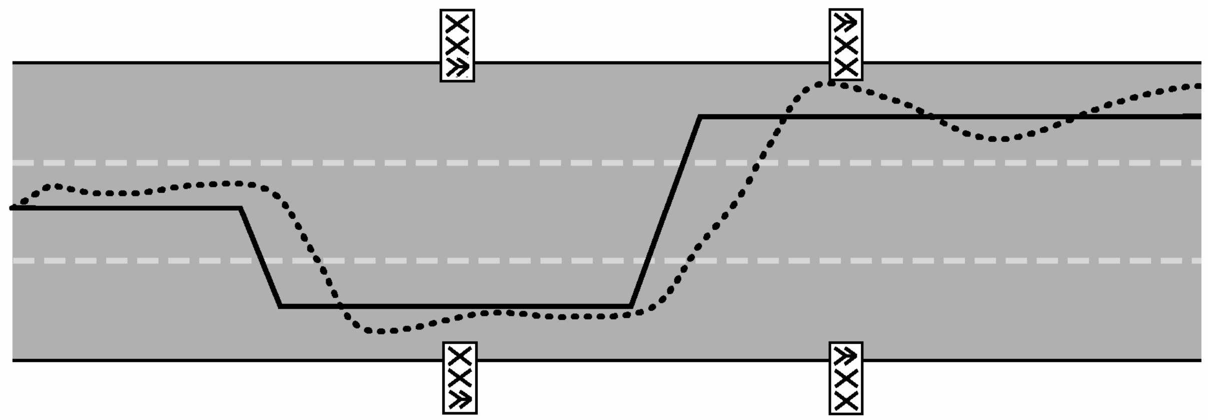

Figure *17*. Example of a normative driving model (solid line) and actual driving data (dotted line) in the Lane Change Test. The area between the normative model and the behavioral data provides an automatic and objective performance measure for driving quality.

Story comprehension.

The answers to the multiple-choice questions about the story sections were evaluated offline. The percentage of correct answers was evaluated for each subject.

Pre-processing of stimulus-response data.

The data of two subjects were missing for the analysis of the secondary tasks. Only 19 of the 21 behavioral data sets could be analyzed. For 1 subject the data was completely missing due to a failure of the recording system. In another session, the response button failed. The sequence of tones in each of the eight 3-min segments used in the simple tone detection task was determined. Similarly, the exact beginning of each keyword in each of the 3-min ongoing story blocks was analyzed in a sound editing program (Adobe Audition v2.0).

Tone detection performance analysis.

For the tone detection task, the responses that occurred between the start and the end of the secondary task were extracted for the analysis and mean response times as well as d′ were calculated based on the stimulus and response markers. Correct answers were counted for reactions that occurred within a given time window of 200 to 3000 ms of stimulus onset. Answers faster than 200 ms were excluded from the analysis. If two answers were given, the first was evaluated and the second answer ignored. The response to one of the two tones could either have been a correct response, a wrong response (i.e. a response to the low tone), or a miss. Reactions to the 500 Hz tone were counted as hits. A missing answer to the 400 Hz tone was counted as a correct rejection. From the total number of high tones presented, the z score of hits, and the z score of misses d′ was calculated for each experimental block. In order to avoid infinite z scores for hit rates of 1.0 or false alarm rates of 0, all hit and false alarm rates were routinely corrected as recommended by Snodgrass and Corwin (1988). According to their procedure, conditional probabilities are calculated by adding the value 0.5 to the hit and false alarm frequency. Each sum is then divided by the number of high tone trials +1 or low tone trials +1 respectively. Reaction times were calculated for every correct answer given within the response time window. As for d′, the mean value for each experimental block will be reported in the results part below.

Word detection performance analysis.

The keyword trials in the word detection task appeared within the different sentences of an ongoing story. In order to calculate d′ for this task, the data of the 180-s story blocks were divided into consecutive 3-s trials starting from the beginning of a story block. No-stimulus trials were created until a keyword fell into one of these 3-s windows. Then a new stimulus trial was created at the beginning of this stimulus. Time windows shorter than the length of 3 s could not be evaluated with this procedure and were therefore eliminated. A total number of 96 trials (98 trials for one subject that heard one story block twice) for the keywords and 234 trials for the no-stimulus trials were obtained over all story blocks resulting into a mean of 18 keywords (range= 17 - 22) for each of the eight story blocks. However, since the keyword material was presented within a story it could be the case that two or more keywords appeared too quickly after one another which made it impossible to assign a participant's response to the appropriate keyword. These overlapping 3-s keyword trials had to be excluded prior to the analysis resulting into a mean of 12 keywords (range = 8 - 16) per block. Generally, responses later than 3 s after the stimulus have been scored as a miss for the trial and counted as a false alarm for the following non-stimulus window. The accuracy measure (d') was calculated in the same manner as in the tone detection task described above. Finally the mean hit reaction times and d' of both tasks were exported for the statistical analysis.

3.2.7.2 Neurophysiological data analysis.

The EEG data was band pass filtered (bandpass: 0.5 – 56.25 Hz), resampled to 125 Hz and re-referenced to the average of the two mastoid electrodes (TP9, TP10). Each recording was visually inspected for muscle artifacts, electrode drifts and other distortions of the data and the respective segments were marked. With the marked artifacts and ECG channels removed, an ICA was run on the individual data. Eye-movement-related independent components were removed based on visual inspection of the topographic distribution and component activation over time. After component subtraction previously rejected data segments were restored and each recorded data set was again visually inspected for additional artifacts that could not be removed by the ICA. Then data of each experimental block were segmented using a sliding window technique (4 s windows, 75 % overlap) before computing the power-density discrete Fourier spectrum (hamming window, padding factor 4 used for interpolation) on these data windows. After this step artifact-prone data segments were eliminated. The 4-s windows were averaged to receive a power density spectrum for each experimental block. To take inter-individual differences in base level of power spectra into account, power values in each channel and condition were divided by the sum of the power from 5 - 40 Hz at the same channel and condition. This moderate frequency window was chosen to minimize the contribution of noise of the very low and high frequencies. The spectral analysis of the data produced 15 sub-bins within a 1 Hz range. To summarize the data into frequency bins of 1 Hz steps, all sub-bins within this Hz range were averaged to obtain a single value. For example, to obtain a single frequency value for the 5 Hz bin, the power data greater or equal 5 Hz and smaller than 6 Hz were averaged.

For the analyses of EEG alpha power, a semiautomatic individual alpha peak adjustment was performed for each high workload condition (word detection with and without LCT). For each subject the power spectra of 11 electrodes (F3, F4, FZ, C3, CZ, C4, P3, PZ, P4, O1, and O2) were plotted and the electrode with the highest power value in the range of 5 - 15 Hz was automatically marked. After visual inspection of the spectral peak, the power with a range of 2 Hz around the individual and task-specific peak was extracted for all eleven electrodes and statistically analyzed. For a few subjects with no identifiable peak a standard alpha frequency of 10 Hz was used.

3.2.7.3 Peripheral physiological data analysis.

In general, only the ECG channel next to the upper end of the sternum was used for the analysis. However, for two subjects the data of the second ECG channel was used due to bad data in the first channel. Epochs according to experimental blocks were built and a standard peak detection algorithm was run over the data to identify R-peaks and the so-called “normal-to-normal” intervals, i.e. all intervals between adjacent QRS complexes in the ECG signal were identified.

The heart rate and the three dispersion measures of heart rate variability were calculated for each block from the recorded ECG signal. The resulting data were checked for outlier values which seemed physiologically implausible and which might be the result of difficulties with the data recording. The plausible intervals were chosen to be (1) 40 - 200 for bpm, (2) 10 - 150 for SDNN, (3) 0 - 100 for PNN50 and (4) 100 - 1000 for RMSSD. Based on an outlier analysis of the heart rate data, 1 subject had to be completely excluded from the further analysis (M_{s02} = 108 bpm, SE_{s02} = 0.91). A second subject (M_{s05} = 99 bpm, SE_{s05} = 0.55 compared to M_{all} = 64 bpm, SE_{all} = 1.97 of all other 18 subjects) was excluded from the heart rate analysis due to outlier values. Mean values of the four measurements for each experimental block and subject were exported for the statistical analysis.

3.2.7.4 Statistical power.

21 subjects were tested in Experiment 1, 18 subjects in Experiment 2, and 30 subjects in Experiment 3. Analyses of statistical power were performed post-hoc and only when necessary for better interpretation of the results. Statistical power can provide information about the probability of rejecting a test's null hypothesis given that it is in fact false, i.e. the probability to find an effect of a specific size if it truly exists in the population (Faul, Erdfelder, Lang, & Buchner, 2007). Calculations of statistical power have been performed using G*Power (v. 3.0.8) and are reported in the respective results parts of each experiment whenever considered helpful for the interpretation of the results.

3.3 Results

All statistical tests reported in this section assumed an α-level of .05. If ANOVAs showed a violated sphericity the Greenhouse-Geisser corrected p-values (pGG) are reported.

3.3.1 Lane Change Test

High cognitive demands from a secondary task had considerable impact on the performance of the lane changes in the driving simulation. As can be seen in *Figure 18*, a higher mean deviation from the reference lane is observed for the difficult word detection task. A two-sided t-test for paired samples showed a significant effect ($t(1, 17) = 9.755$, $p < .001$, $\eta^2 = .85$). Means of the standard deviation of the steering angle did not show any significant differences in a two-sided t-test for paired samples. However, a post hoc statistical power analysis ($\alpha = .05$, $n = 18$, effect size $dz = .356$, $r = .944$) revealed a relatively low statistical power ($1 - \beta = .423$) for this test.

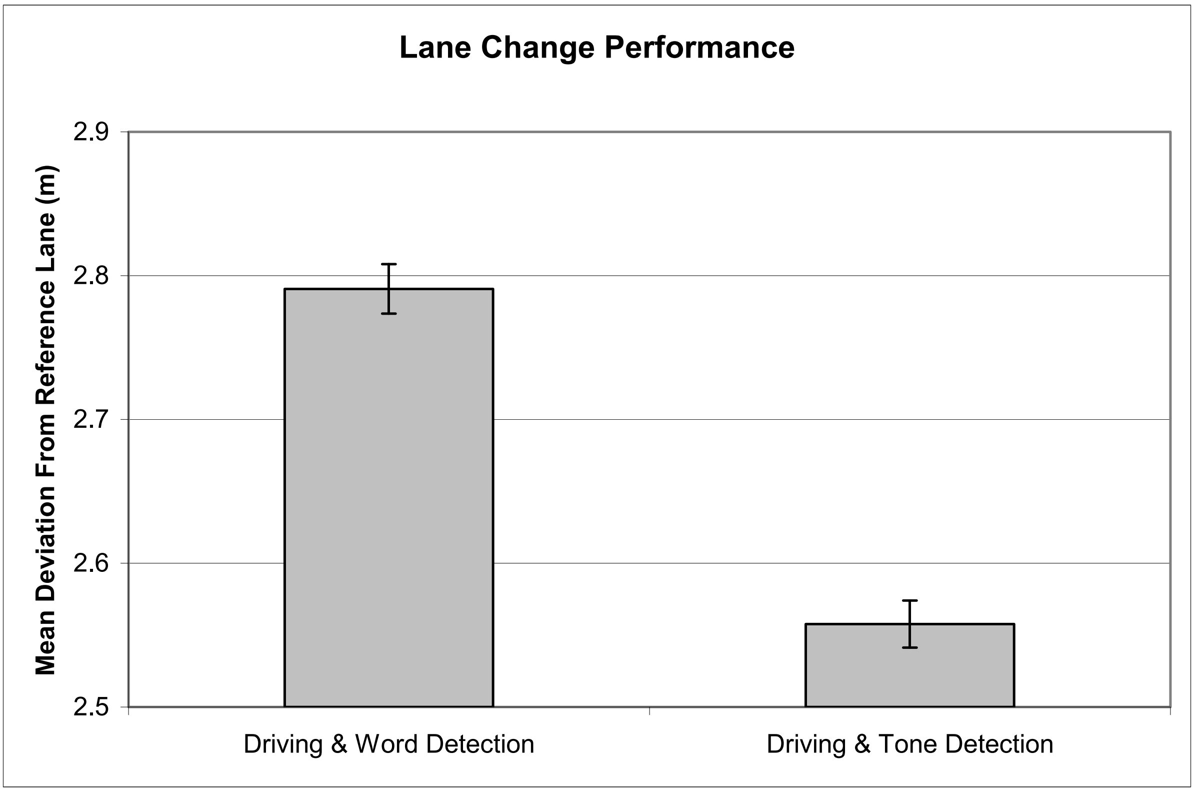

Figure 18. Mean deviation (n = 18) from a standard driving course model when subjects either performed a highly loading word detection task or an easy tone detection task in parallel. Error bars represent the standard errors.

3.3.2 Recall of Story Content and Detection Performance

Two subjects provided no data due to a failure of the recording equipment resulting into a data analysis over 19 subjects. First, the answers to the questions about the story content were analyzed. The 19 subjects showed an overall high performance of 92.11 % correct answers. No systematic errors could be identified for any particular question. A total of 8 questions were asked to each subject. Four questions referred to text sections that were presented in parallel to the LCT. No difference was observed between questions to texts presented with or without LCT, i.e. the LCT and the story listening task did not interfere with each other. This indicates that participants put the same priority on the listening task regardless of driving. Order of sessions, i.e. the presentation of the LCT in the first or the second half, didn't have an effect on the percentage of correct answers either (driving and word detection presented in the first half = 92.5 % correct answers; driving and word detection presented in the second half = 93.61 % correct answers; word detection without driving presented in the first half = 93.33 % correct answers; word detection without driving presented in the second half = 90 % correct answers). These data serve as a control by showing that all subjects listened carefully to the content of the story.

The mean reaction time and mean d′ over blocks for the responses to the auditory detection tasks are shown in *Figure 19* and *Figure 20*.

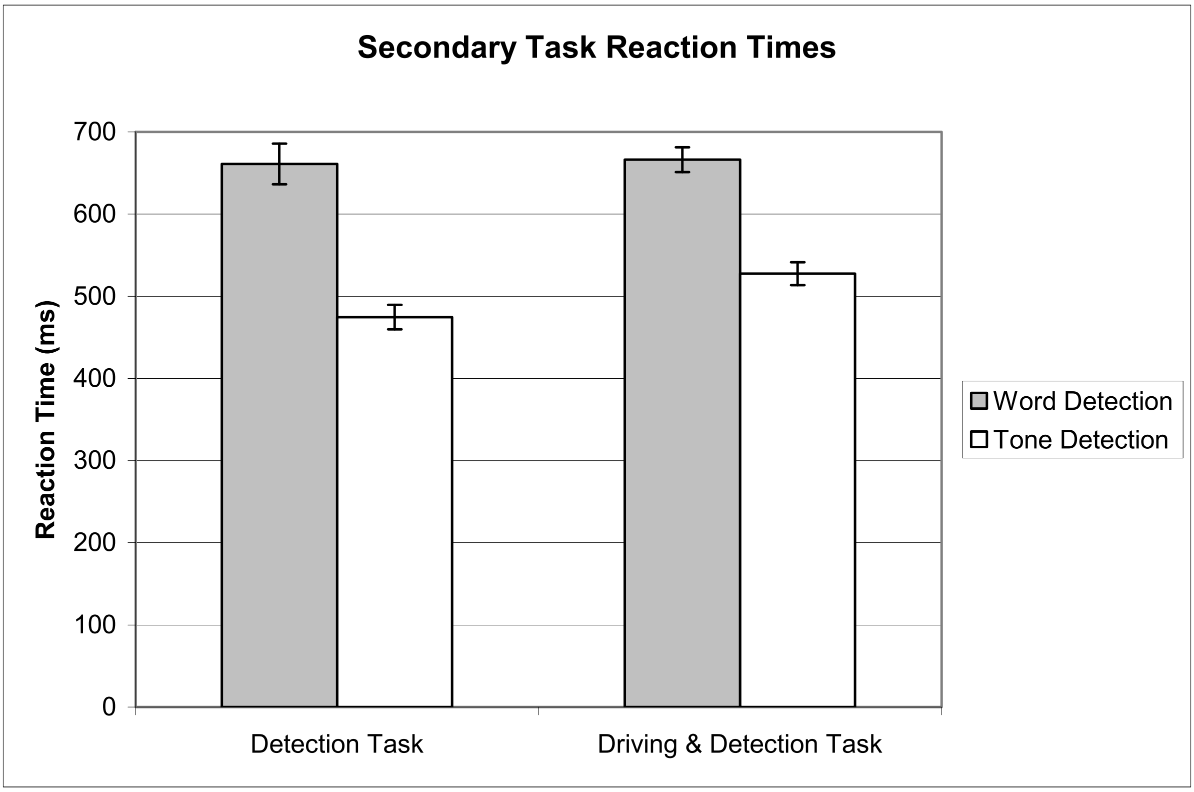

Figure 19. Mean reaction times (n = 19) for the detection of words or tones in the secondary task with or without driving in parallel. Error bars represent the standard errors.

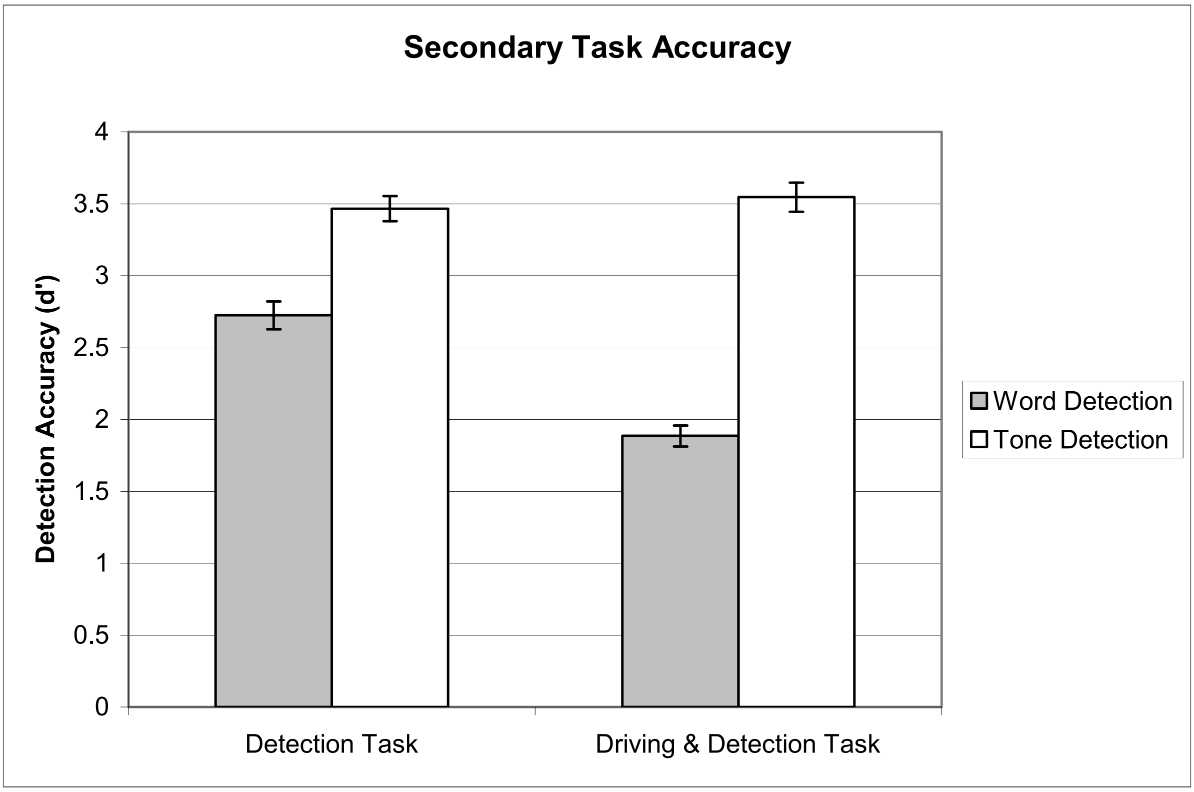

Figure 20. Mean d′ values (n = 19) for the detection of words or tones in the secondary task with or without driving in parallel. Error bars represent the standard errors.

The more difficult word detection task produced longer reaction times and worse detection accuracies than the easy tone detection task. While performing the LCT in parallel had no effect on reaction times, detection accuracy was worse for the word detection task. A two-factorial repeated-measures ANOVA with the factors "number of tasks" (detection task vs. driving and detection task) and "task difficulty" (easy tone detection vs. difficult word detection) revealed a significant main effect of task difficulty for both performance measurements (reaction times: $F(1, 18) = 26.881$, $p < .001$, $\eta^2 = .599$; detection accuracy d′: $F(1, 18) = 88.873$, $p < .001$, $\eta^2 = .832$). For the number of tasks factor reaction times did not show any significant main effect ($F(1, 18) = 1.452$). For d′ a clear tendency towards a higher performance in the single task condition could be observed that just failed to reach the α-significance level of .05 for the number of tasks factor ($F(1, 18) = 4.010$, $p = .061$, $1 - \beta = .78$ for $\alpha = .05$, $n = 19$, $\eta^2 = .182$, $r = -.001$). However, the interaction between the two main factors was significant for d' ($F(1, 18) = 4.895$, $p < .05$, $\eta^2 = .214$), but not for the reaction times ($F(1, 18) = 2.262$).

3.3.3 EEG Alpha Power

Two sets of statistical analyses were run for the 11 EEG scalp electrodes: one for the three midline electrodes (FZ, CZ, PZ) and one for the lateral electrodes (F3, F4, C3, C4, P3, P4, O1 and O2). For the latter type of analysis, a comparison between left and right hemisphere electrodes was initially run to assess any differences in the lateralization of the effect. If no significant difference was observed, the values of left and right electrodes were always averaged to a single mean for the subsequent analyses. If the two sets of midline and lateral electrode analyses did not differ in their results, then only the statistical parameters of the lateral electrode analyses were reported in the following results part. Greenhouse Geisser corrected values were reported in cases of violated sphericity.

3.3.3.1 Individual alpha peak adjustment.

Attempts to use the alpha peak obtained from the 10-min baseline recording with eyes closed as a reference for individually defining the alpha power range in the two high workload conditions were not successful. Alpha power peaks recorded during different task conditions were not comparable within a subject, i.e. the alpha peak recorded in the baseline condition with eyes closed was qualitatively different from alpha power recorded under task demands (see example in *Figure 21*). This qualitative peak difference can be seen for the frequency range, the power amplitudes, and the electrode where the maximum occurred.

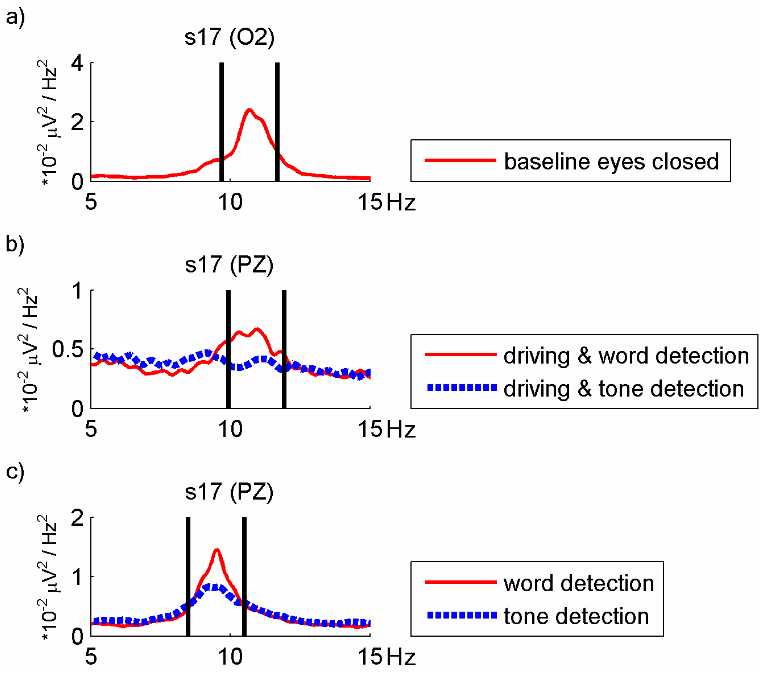

Figure 21. Individual and task-specific EEG alpha peak adjustment procedure for 1 subject in Experiment 1 as it has been performed for all three experiments in this thesis. a) Baseline recording with eyes closed, b) detection tasks without driving and c) detection tasks with simultaneous Lane Change Test. The two vertical lines signify the 2 Hz frequency range around the amplitude peak that has been extracted for subsequent alpha power analyses. Please note that y-axes have different scaling.

An univariate ANOVA was calculated to compare individual alpha peak frequency values across 20 subjects. The analysis included the within-subject factor "task type" (baseline with eyes closed vs. word detection vs. driving and word detection), but it revealed no significant difference between experimental conditions ($F(1.408, 26.755) = 1.374$). A post hoc power analysis ($\alpha = .05$, $n = 20$, $\eta^2 = .067$, $r = .647$) in G*Power revealed a relatively high statistical power ($1 - \beta = .841$) supporting the observation that peak frequencies did in fact not differ between tasks.

However, a comparison of individual alpha power amplitudes supported observations of qualitative differences in alpha peak between the conditions. Across all subjects ($n = 20$) the mean alpha amplitude for the baseline is much higher than the amplitude in the two demanding task conditions (see *Figure 22*).

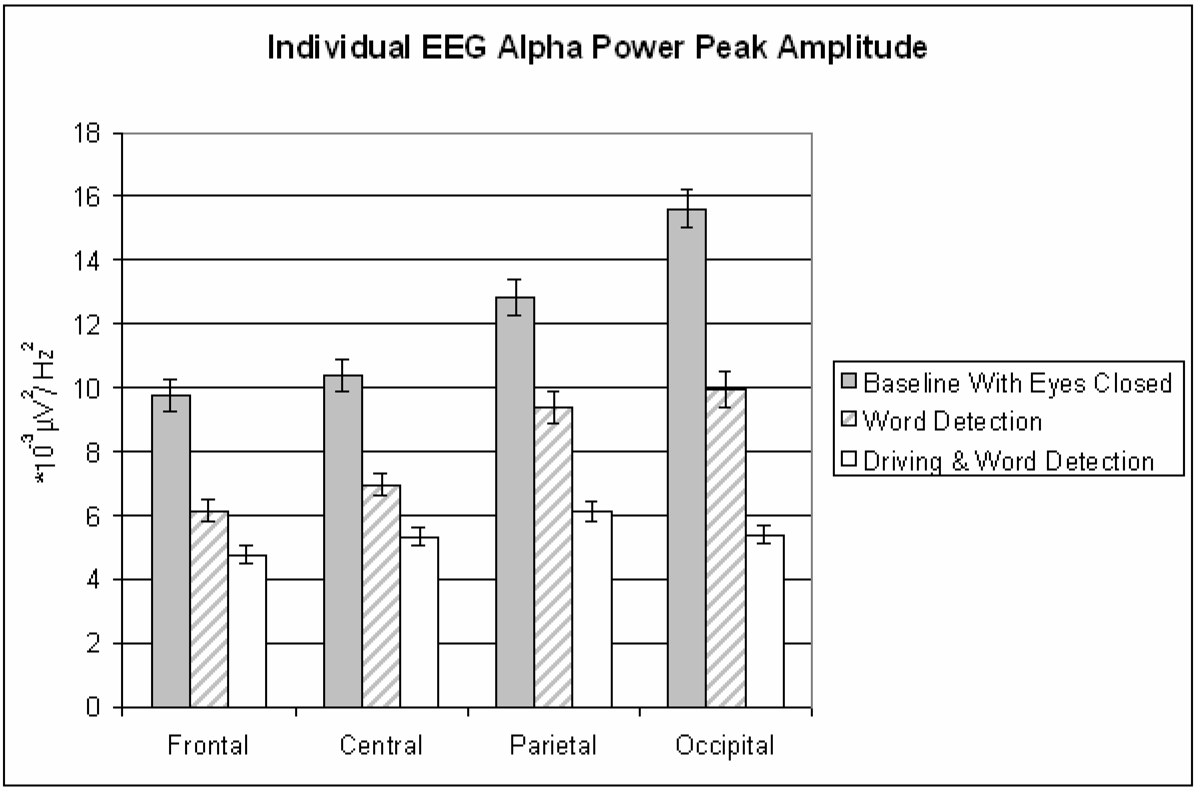

Figure 22. Comparison of individual EEG alpha power recorded during the baseline with eyes closed, word detection and word detection with Lane Change Test. Error bars represent the standard errors.

Highest power values were recorded at occipital electrodes during the baseline condition, followed by the word detection condition with highest values and a similar distribution. The lowest power values were obtained during the word detection with simultaneous driving. In this condition the maximum effect was over parietal electrodes. A repeated-measures ANOVA with the two main factors "electrode location" (frontal vs. central vs. parietal vs. occipital) and "task type" (baseline with eyes closed vs. word detection vs. driving and word detection) over lateral electrodes showed significant effects for the electrode location ($F(1.6, 30.404) = 21.384$, $pGG < .001$, $\eta^2 = .530$) and task type factor ($F(1.498, 28.468) = 39.747$, $pGG < .001$, $\eta^2 = .677$) as well as a significant interaction between factors ($F(2.591, 49.236) = 14.489$, $pGG < .001$, $\eta^2 = .433$). Helmert contrasts indicated that alpha power was significantly different in all three task conditions (baseline vs. two word detection conditions: $F(1, 19) = 42.58$, $pGG < .001$, $\eta^2 = .691$; word detection vs.

driving and word detection: $F(1, 19) = 29{,}35$, $pGG < .001$, $\eta^2 = .607$). The significant interaction stemmed from a stronger pronounced effect over occipital electrodes. Similar results were obtained from the analysis of the three midline electrodes.

3.3.3.2 Individual EEG alpha power effects of mental workload.

By following the adjustment procedure described in the preceding section, the individual alpha frequency power was extracted and the power values were averaged over subjects. The means over subjects for the workload-relevant conditions and the distribution of the effects over the scalp can be seen in *Figure 23*.

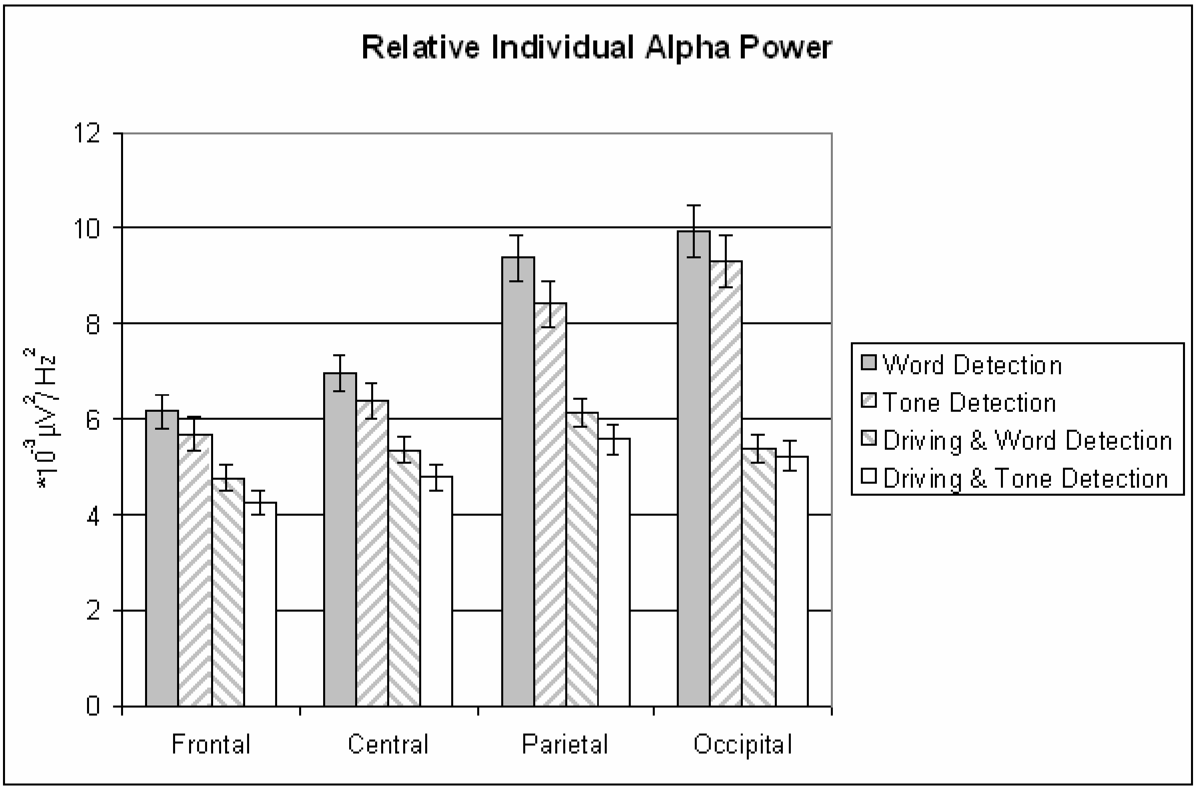

Figure 23. Mean individual EEG alpha power over all subjects ($n = 20$) observed under different levels of mental workload and recorded either when subjects simultaneously performed the Lane Change Test or when they performed the cognitive task alone. Error bars represent the standard errors.

A repeated-measures ANOVA with the three main factors “number of tasks” (word detection vs. driving and word detection), “task difficulty” (difficult word detection vs. easy tone detection) and “electrode location” (frontal vs. central vs. parietal vs. occipital) was run. As the analysis showed, all main effects were significant. The demanding word detection task conditions elicited higher power amplitude values than the simple tone detection tasks ($F(1, 19) = 9.024$, $pGG < .01$, $\eta^2 = .322$). When comparing alpha in the two experimental conditions in which subjects were simultaneously driving with the conditions when they performed no parallel task an attenuation of alpha power can be observed ($F(1, 19) = 32.604$, $pGG < .001$, $\eta^2 = .632$). In the conditions without driving alpha power increased from frontal to occipital electrode sites ($F(1.489, 28.298) = 15.12$, $pGG < .01$, $\eta^2 = .443$). The interaction between the task difficulty factor and the electrode location factor was not significant ($F(1.573, 29.885) = 1.665$) which speaks for even distribution of the effect over all electrodes. Moreover, the effect is stable regardless if subjects performed an additional driving simulation or whether the detection tasks were performed alone as indicated by the nonsignificant interaction between the task difficulty factor and the number of tasks factor ($F < 1$). The effect for number of tasks, i.e. the reduction in alpha power when performing an additional visuo-perceptive task, is strongest over occipital electrode sites. A gradual alpha synchronization can be observed from the front to the back of the head in the single task conditions which is clearly less pronounced for the conditions with simultaneous driving. This observation is confirmed by a statistically significant interaction between the number of tasks factor and the electrode location factor ($F(1.48, 28.129) = 13.647$, $pGG < .01$, $\eta^2 = .418$).

In sum, a more difficult detection task led to higher individual alpha power over all electrode sites. On the other hand, an additional task with a strong visuo-perceptual emphasis like a driving simulation led to a reduction in the alpha band that is apparent in both detection tasks and strongest at occipital electrode sites. Exactly the same pattern of results was obtained for the analysis of the three midline electrodes.

3.3.4 ECG Indices of Mental Workload

Mean values over all four blocks were calculated for the heart rate and each of the three heart rate variability dispersion measures. All mean variables were tested using the Kolmogorov-Smirnov test for normal distribution. The results showed normally distributed mean values for all ECG variables. The ECG measures were averaged over all four blocks. The means ($n = 17$) for the standard deviation of N-N intervals (SDNN) are displayed in *Figure 24*. The SDNN was decreased when subjects performed the more difficult word detection task regardless of whether they performed a driving simulation in parallel or not. The measure showed higher heart rate variability for the more difficult word detection task compared to the easy tone detection. An effect of the multiple-task situation could not be identified. The means were entered into a repeated-measures ANOVA with the factors "number of tasks" (detection task vs. driving and detection task) and "task difficulty" (simple tone detection vs. difficult word detection). Only the task difficulty factor revealed a significant effect for the SDNN ($F(1, 16) = 4.958$, $p < .05$, $\eta^2 = .237$) while the number of tasks factor failed to reach statistical significance ($F < 1$). The interaction between main factors was not significant ($F(1, 17) = 1.639$) indicating that the task difficulty effect was equally present in both detection task conditions, i.e. with or without simultaneous driving. These results indicate that the more demanding word detection task leads to a decrease in heart rate variability which is in line with reports from the literature.

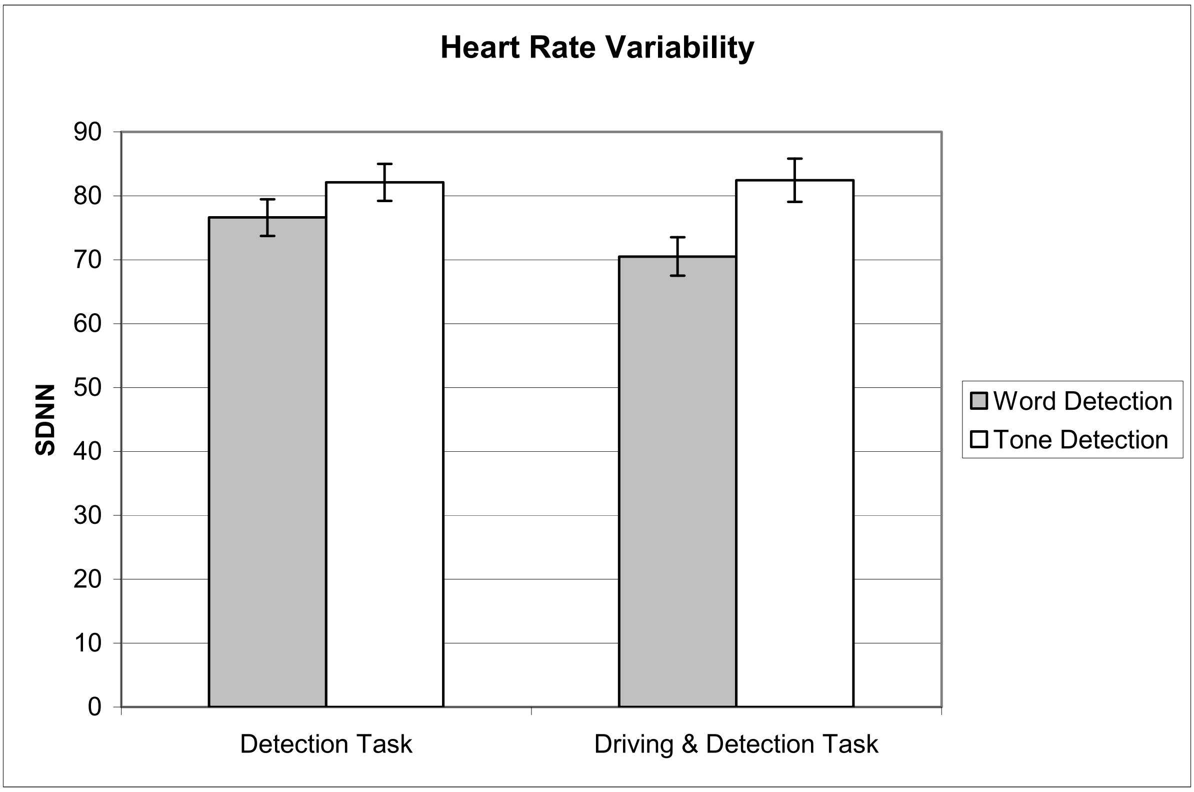

Figure 24. Mean standard deviation of N-N intervals (n = 17) in the ECG for the experimental manipulation of mental workload recorded either when subjects performed the Lane Change Test in parallel or when they performed the cognitive task alone. Error bars represent the standard errors.

The analysis of the PNN50 did not show any significant effects in a repeated-measures ANOVA with the same two main factors task difficulty ($F < 1$) and number of tasks ($F < 1$). The same ANOVA for RMSSD also revealed no statistically significant results for the task difficulty factor ($F(1, 15) = 1.427$) and the number of tasks factor ($F(1, 15) < 1$, $\eta^2 = 0$).

The mean heart rate (n = 18) for the four experimental conditions is shown in *Figure 25*. The results show an increase in heart rate for the easier tone detection task and a ANOVA similar to the analyses reported for the other measures revealed a significant effect for the task difficulty main factor ($F(1, 17) = 4.532$, $p < .05$, $\eta^2 = .21$) while the number of tasks main factor was not significant ($F < 1$). The interaction between main factors was not significant ($F < 1$).The direction of the effect for the heart rate is against predictions and shall be further discussed in the discussion section.

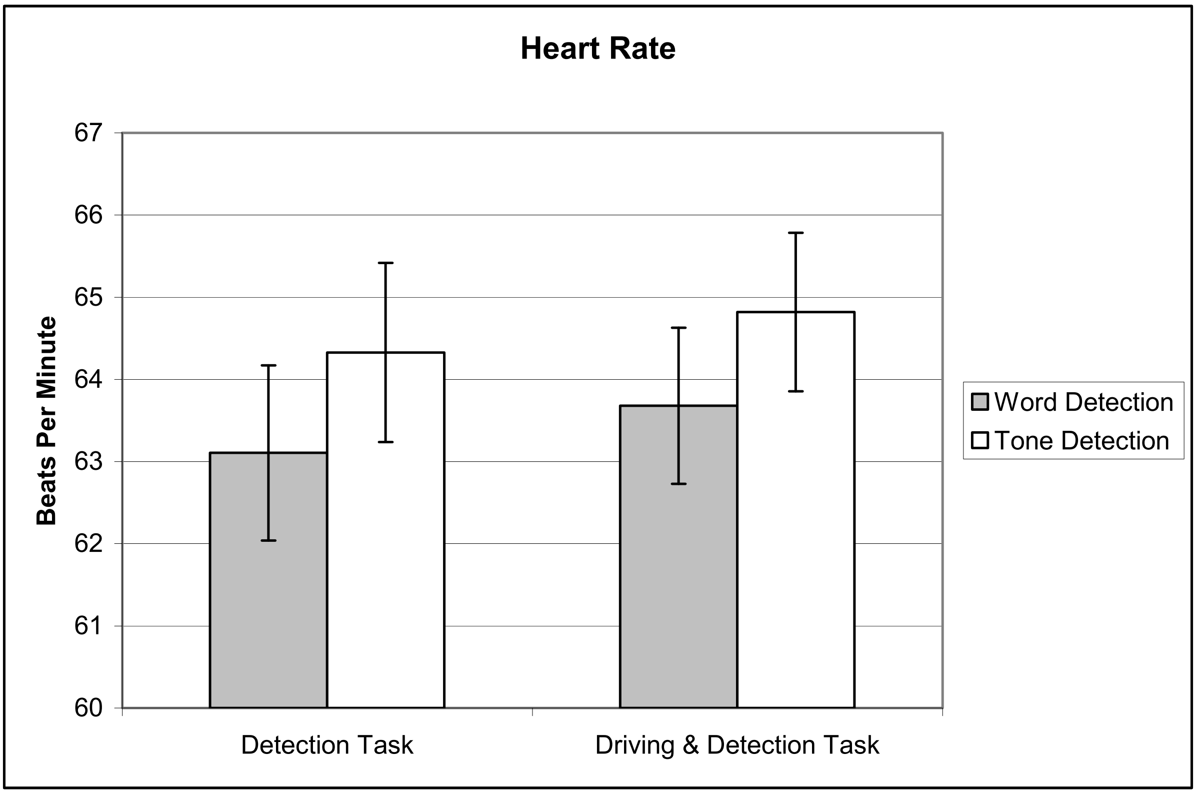

Figure 25. Mean heart rate (n = 18) for the experimental manipulation of mental workload recorded either when subjects performed the Lane Change Test in parallel or when they performed the cognitive task alone. Error bars represent the standard errors.

3.4 Discussion

It has been the aim of this first laboratory experiment to investigate the impact of two different levels of a cognitively demanding task onto the cognitive state of a driver in a driving simulation. Based on the results, it can be concluded that subjects who intensively listen to a laboratory task show a state in their EEG alpha band that was also observed in other laboratory studies during internalized attention (e.g. Cooper, Croft et al., 2003). If compared to terms used in traffic research the observed effect most likely reflects a state described as "endogenous distraction" (Recarte & Nunes, 2003) which could probably also result in "inattentional blindness" (Strayer & Drews, 2007) during a traffic situation. Moreover, a performance decline in the driving and detection tasks reflects the consequences of mental workload. In sum, the results in the EEG alpha band and performance data show that these measurement techniques allow the reliable registration of mental workload even in complex task situations. It can thus be expected that similar phenomena are observed when measuring in real traffic. Only the ECG data delivered contradictory results. In the following the possible implications based on the findings from the experiment are discussed.

3.4.1 Implications Based on the Laboratory Results

This section starts with a discussion of the most important results first, i.e. an interpretation of the EEG alpha band data. In 2002, Recarte and Nunes already noted: "It is difficult to predict which verbal messages might or might not be a source of distraction without empirical testing" (p. 134). Nevertheless, the present results demonstrate that the used tasks successfully manipulated the subjects' workload. Subjects who simply detected high tones among low tones show a lower alpha power level. On the other hand, a clear increase in the EEG alpha band is observed when subjects put much effort into listening to the content of a story and simultaneously trying to confirm the appearance of certain words. The different difficulty levels of the cognitive tasks have a clear impact on the EEG alpha frequency spectrum. As presented in the theoretical part of this work, the results of earlier works indicated that alpha power may be related to the inhibition of task irrelevant cortical areas. In this context, many authors suggested a ring of inhibition that surrounds the active processing area and that shows increased alpha oscillations (Cooper, Croft et al., 2003; Pfurtscheller, 2003). It has been assumed that this phenomenon mirrors a top-down process in order to focus on a task. The observation of an alpha power increase during a straining listening task implies that similar processes take place here. However, it has to be admitted that with the limited number of electrodes used in the present analysis a ring of inhibition can not be confirmed, but a broadly distributed effect over the whole head indicates a global inhibition mechanism that obviously serves high concentration.

Even in a situation in which the subjects were primarily involved in the driving task, the impact of mental workload induced by a secondary task can clearly be shown. Other authors demonstrated in similar experiments that alpha oscillations were eliminated under multiple-task conditions (Fournier, Wilson, & Swain, 1999; Gundel & Wilson, 1992). They referred to an increase in visual stimulation to explain the observed effect. In fact, there is also a significant increase in visual stimulation during driving in the experiment at hand. And indeed, a general attenuation of the general alpha power level under multiple-task conditions can be observed which is furthermore increasing from frontal to occipital electrode sites compared to the single-task condition. Thus, the attenuation is maximal over a cortical area where visual processing is predominantly taking place. Despite this influence on the alpha band a diametrical difficulty effect can be observed at the same time. In contrast to other works this may be a possible reason why no floor effect is visible in the EEG here. It can be hypothesized that the required focusing or the inhibition of distractors due to the additional task lead to an initial increase in the task difficulty effect and that this could be the reason for the stability of the effect. In this way, the results are consistent with reports about endogenous distraction during conversations while driving. It may be possible that an active suppression of seemingly task-irrelevant stimuli leads to a downright "inattentional blindness" as it has been demonstrated in the case of cell phone conversations in driving experiments (Strayer & Drews, 2007). With the present results a first-time demonstration of electrophysiological correlates in the alpha band of endogenous distraction during driving has been achieved.

Some aspects concerning the EEG analysis shall be considered in the following. The EEG analysis was performed individually and task-specific by determining the peak-locked alpha frequency range. From previous works on the EEG alpha band (e.g. Klimesch et al., 1996) it is known that there are great inter-individual differences in the range of frequency bands which require an individual adjustment. Typically the peak in the alpha band is used for this adjustment. The initial idea was to use a narrow 2 Hz frequency range around the individual alpha peak recorded in the baseline condition with eyes closed. The alpha power in this condition usually shows high amplitude and a clearly defined peak. But after inspecting the actual peaks in the alpha spectrum for each subject in each alpha-relevant task condition (baseline and two high workload conditions, see appendix A) this approach had to be rejected. The direct comparison reveals that the qualitative differences between the tasks are reflected in most of the individual EEG power spectra. The use of only one individual alpha peak for all conditions appeared to be unreasonable and therefore task-specific power values around the alpha peak were extracted for each subject. In order to be able to compare the results the same analysis procedure was used for all three experiments. However, the effects in the laboratory are so pronounced that even a standard alpha band analysis in the range of 8 - 12 Hz produces significant results (see appendix B).

Another important point of discussion concerns the time-consuming offline artifact handling through visual inspection which always implies a subjective bias. However, with respect to the challenging data situation in the real-traffic driving situation it still remains the first method of choice. As stated above, the analysis was similarly performed to guarantee that a comparison of the results is possible. With the development of automated methods and their increasing availability in commercial software packages this very resource-intensive type of artifact handling will hopefully be dispensable in future investigations.

Before starting a general discussion that concerns all data sources, the remaining results are briefly discussed here and possible conclusion are drawn. The analysis of the drivers' performance in the LCT reveals that under high secondary task demands drivers deviate to a higher degree from an ideal reference model than when task load is low. These results speak for a successful mental workload manipulation and they stand in line with previous works showing the sensitivity of the LCT to drivers' mental workload (e.g. Mattes, 2003). This outcome also provides support for the concept of a central bottleneck of attention (Levy et al., 2006), i.e. that driving is not completely independent from an auditory secondary task, but that both tasks address limited attentional resources at a central processing stage. However, a second performance measure, the standard deviation of the steering angle fails to differentiate between the workload conditions. The lack of an effect may be explainable when considering that subjects only used the arrow buttons on the keyboard for driving instead of a steering wheel and foot pedals. This task setup was an artificial restriction through which the task lost some of its face validity and for the subjects this might have resulted in lower task-involvement. It seems likely that steering behavior was not an accurate measure under these conditions. Nevertheless, it has been considered of higher importance to minimize the proportion of artifacts from muscle activity than to achieve an absolute maximum of realism for the driving simulation inside the laboratory. Since on-road experiments were planned as well this trade-off seemed acceptable. Moreover, it has to be considered that performing the LCT via keyboard buttons at the right hand and performing the detection task via button press with the left index finger might have led to inferences. For future experiments it may be important to choose better distinguishable input devices (e.g. buttons attached to index fingers).

Overall, the use of a LCT has been ideal for the investigation of car driving in the laboratory. The task delivered a constant measure for driving quality and it strained the driver on a constant level. Task demands were given by a continuous optic flow combined with a constant regulatory task for the driver. Moreover, at least one measure of the LCT has been sensitive to the workload manipulation by a secondary cognitively loading task.

Besides the performance in the primary task, secondary task detection performance also shows that the combined story listening and word detection task is more difficult than the simple differentiation of high tones from low tones. When detecting words in a story, subjects needed more time to react and they were less accurate in their responses. For the reaction time measure this effect of task difficulty is stable across the number of tasks performed in parallel, i.e. regardless whether the subjects simultaneously performed the LCT or not. However, the d' accuracy data show that the difference between the word and tone detection is more pronounced when performing a simultaneous LCT than without it. This result is according to the hypothesis, i.e. it reflects the impact of additional workload by the extra task. A high percentage of correct answers to the questions about story contents proofed that subjects really listened carefully to the stories. It can thus be concluded that subjects were truly in a state of high mental workload during this part of the experiment.

The results of the ECG can not be unambiguously interpreted. Four measures of the ECG were used to assess effects of mental workload. Only one measure of heart rate variability (SDNN) is significantly reduced under high workload conditions and this effect is stable in the single-task and the multiple-task condition. The direction of the effect is in line with observations reported in the literature. However, the heart rate shows an increase under the easier task conditions. This result seems contrary to the expected outcome. High mental effort is usually reflected by an increase in heart rate. A suitable explanation for the divergent finding is hard to find. Certainly, it has to be considered that the ECG recording setup of all three experiments of this thesis had a clear disadvantage: No control measure of respiration was included. Respiration might have had an important impact on heart rate and heart rate variability and usually the signal needed to be corrected for this influence. For this reason, the ECG results of the three experiments in this thesis need to be interpreted in consideration of this limitation. Nevertheless, even with the missing control measure the present pattern of results obtained here could not have been explained.

A more general concern relates to the missing of another experimental condition. It might have been a good idea to include a condition in which subjects only had to drive. With the data from such an additional condition the pure influence of the driving task onto the EEG could have been examined in more detail. In the present investigation the data could only be compared to a resting baseline with eyes closed which can certainly not account for visual influences. Speculations regarding contradictory effects during the multiple-task condition (mental workload effect vs. effect of visual stimulation) could have been closer investigated with such an extra condition.

A general aspect concerns the fact that different mental states were induced for a time period of 3 min. It has to be considered that such a relatively long period of time may have produced a certain blur in the data, i.e. it was difficult to control whether subjects were indeed always completely involved in the task. It could have been possible that for a few subjects confounding influences due to perceived monotony in the tone detection task were produced and recorded in the EEG. Of course such an influence would also always depend on personal factors to a certain degree. Although the tasks included a continuous performance control it could not be expected that subjects were in exactly the cognitive state that had been predicted. In fact, a perfect task involvement can only be reliably guaranteed by running a field experiment.

3.4.2 Conclusions for the Real-Traffic Driving Experiment

This final section is concerned with implications that played a role in the planning of the subsequent real traffic experiment. An argument that may apply to every secondary task paradigm concerns the prioritization of tasks. If the subjects really primarily attended to the quality of their lane changes or whether they cared more to achieve a fast and accurate detection performance was difficult to control. Nevertheless, this presented only a minor concern for the present investigation, since no classical loading task paradigm (see section 2.3.3) has been used. An autonomous shift of priorities would have only been problematic if for example this led to a complete concentration on a single task by neglecting the other task. However, a constantly high number of correct answers to the contents of the texts confirms a persistently high amount of attention to the secondary tasks.

It has been the intention to design both workload conditions of the auditory secondary task as similar as possible. Except for the text processing component which made the task more difficult, this seems to have been the case. However, a perfect comparability of both detection performances could not be assured with the design at hand. The detection of words is different from the detection of single tones.

At this point, the introduction of an extra, continuous stimulus-response task in the second field experiment seemed justified. Such a task guarantees a direct comparability between sections of high and low workload. For the field experiment it has been refused to include different types of tasks for high and low workload. Instead a highly demanding cognitive task was compared to sections of pure driving. Moreover, a second cognitively demanding task was added to the design.

Another advantage resulted from the use of a constant additional task across workload conditions. Such a task enables the investigation of event-related potentials which provide an extra electrophysiological measure of brain activity. Phenomena like "inattentional blindness" can indeed only be examined if a design is used that requires the response to an additional stimulus in the driving task. An extra stimulus embedded into the driving task means a lot of extra effort for the experimental procedure (e.g. designing a car following task). Even in this case, the introduction of another constant stimulus reaction task seems the better idea, since it allows the examination of potential consequences of mental workload in more detail.

It is certainly the case that in a real-traffic investigation a situation is observed that includes an almost unconfined number of degrees of freedom for driver actions. Thorough planning and instructions for the drivers have to create an experimental setting that allows a controlled manipulation of mental workload. However, a complete control can never be achieved in a field experiment.

Finally, it remains to be demonstrated for the real-traffic experiment that a measurement technique like the EEG which is well-known to be vulnerable to artifacts can provide results analog to the results in the laboratory experiment. In other words, it is expected that an increase in alpha power can be observed when performing a cognitively demanding listening task in parallel to driving which indicates active inhibition processes in the brain. Foregoing real-traffic experiments are rare, but the few investigations that exist (e.g. Schmidt, 2006) and the apparent effects from the laboratory leave ample room for optimism.

4 Experiment 2

4.1 Aim and Hypotheses

The previously described experiment investigated the influence of a cognitively demanding task presented in parallel to a simulated driving task. The results demonstrated that amplified alpha oscillations in the human electroencephalogram may serve as an indicator for internally directed attention under increased information processing demands. However, it has been argued that the artificial situation inside a laboratory can hardly reflect the complex situation in real traffic (cf. Recartes & Nunes, 2003). Therefore, this driving experiment was designed to mimic cognitively demanding situations for the driver in a maximal realistic environment. As in the previous experiment, measures from multiple sources were recorded and analyzed. Besides the auditory content monitoring task, a mental arithmetic task was included in the design. This time the requirement to respond to a stimulus was omitted for the secondary task, but the difficulty level of the auditory monitoring task was increased by mixing the story with distracting radio news. Both tasks were supposed to address central information processing capacities directly. Moreover, a continuous tertiary auditory stimulus-response task was added to obtain a constant measurement of cognitive spare capacities throughout the experiment. This experimental design is similar to a true subsidiary secondary task design (see section 2.3.3).

It has been hypothesized that the drivers' high mental workload induced by either of the two straining secondary tasks would be observable in each of the various measures and that the results of the laboratory experiment would thus be replicated on road. Similar to Experiment 1, the drivers' EEG was expected to show an enhancement of alpha power during periods of high workload, i.e. during demanding secondary task performance as opposed to just driving without additional task. Behavioral performance to the tertiary stimulus-response task should be diminished when the driver is under the influence of high workload. Moreover, event-related potentials to the stimuli of the tertiary task should show characteristic workload-related changes, i.e. a decrease of the P300 amplitude as well as an increase in the amplitude of the N100 due to enhanced attention-intensive processing. Although possibilities for the interpretation of ECG is limited without a respiration control measure (see Experiment 1) it was hypothesized that heart rate should be increased under periods of high workload and that the dispersion measures of heart rate variability were expected to be decreased under high workload.

Although the experiment took place in an open environment that involves uncontrollable confounding factors like weather and other road users, it was tried to minimize the influence of the road traffic through careful instructions of the drivers, e.g. drivers had to drive at a constant speed, stay on the right lane, and only pass other cars if this was inevitable.

4.2 Methods

4.2.1 Subjects

Eighteen employees of DaimlerChrysler were recruited through a general announcement requesting participation in an experimental study. There were 5 females and 13 males ranging in age from 19 – 32 with an average age of 26. For participation in the study, subjects were initially required to be right-handed, possess a valid drivers license, have driven for six years, be a regular driver on a weekly basis at least, speak German as their first language, and have no health conditions that might present a risk to them safely completing the study or could potentially confound the results (e.g. hearing problems, psychotropic drug medication). When asking participants on the test day, one subject admitted to be left-handed and four subjects had less than six years of driving experience (1 subject 4 and 2 subjects five years). Two of the more experienced drivers drove less than weekly. No information about driving behavior was available for 4 subjects. After participation in the experiment, every subject received a gift from the Mercedes-Benz accessory shop as compensation for their efforts.

4.2.2 Experimental Setup

4.2.2.1 Test Course

The chosen route was a two-lane German highway (B10 between Esslingen am Neckar and Göppingen, see *Figure 64* in appendix C) with an overall constant speed limit of 100 km/h and a maximum length of approximately 30 km for a one way drive. Drivers were instructed to drive relaxed and to avoid driving maneuvers that were not absolutely necessary (i.e. stay on the right lane and only pass a car when necessary to maintain the own speed). These instructions were introduced to overcome some of the unwanted variability imposed by the natural traffic environment. The turn point was selected dynamically and announced at the end of the stimulus presentation by the instructor in the backseat of the car. The choice of the turn point depended on traffic, weather, and driving style. Each driver drove two rounds.

4.2.2.2 Experimental Equipment

The experiment was run in a Mercedes-Benz S 320 test-vehicle modified to accommodate various tests of prototype technologies (see *Figure 26*). It was equipped with a computer-generated driver display (VAPS, Engenuity Technologies Inc., 2004) showing a standard driver instrument cluster. The vehicle also included an auxiliary Controller-Area-Network bus (CAN-bus) enabling data to be exported from the vehicles standard CAN-buses to auxiliary computer systems employed to control presentation of experimental stimuli and record data from subjects. 220 V were available inside the cabin to provide power for the equipment. A notebook (standard IBM, 1.4 GHz Intel mobile processor) had been safely installed on a stand above the center-tunnel in the back of the car to enable an easy operation for the instructor. The notebook was used to run the Brainvision Video Recorder software which recorded EEG and video simultaneously. Portable BRAINAMP-amplifiers were installed below the armrest in the back of the car together with the rest of the EEG equipment. The same notebook was used to run the stimulus presentation software (Willmann, 2003) for the continuous tertiary auditory reaction task and to manually write markers into both the EEG and the CAN-bus.

Figure 26. Mercedes Benz S-class (photo copyright 2008 by the Daimler AG).

A video camera was attached to the sun blind next to the co-driver's seat from where it recorded the driving situation as well as any sound from inside the cabin. The auditory stimuli were presented via a loud speaker that was secured on the co-driver's seat. A portable CD-player was used for stimulus presentation of the secondary listening task and a stop watch was used for accurate timing of the experimental blocks. A reaction button was taped to each of the driver's index fingers. Wires leading from the buttons to the parallel port of the EEG notebook were securely taped to the driver's clothes. Each button press wrote a response marker into the marker channel of the EEG.

4.2.3 Physiological Data Acquisition

The EEG data recording setup was similar to the setup described in the previous experiment except that the data was recorded at a sampling rate of 2000 Hz (for more details please refer to section 3.2.3 of Experiment 1).

4.2.4 Task Description and Stimulus Material

The subject's primary task was to drive a predefined route on a German highway at a constant speed of 100 km/h and to autonomously adjust the speed to a lower speed limit where this was required by road, weather, or traffic circumstances. The driver was instructed to stay on the right lane and to avoid any unnecessary passing of cars. In addition, the subject was instructed not to talk during the experiment and to avoid unnecessary movements not related to the driving, since this might have introduced artifacts to the EEG recording. While driving, drivers performed secondary tasks that induced specific types of cognitive load. Secondary tasks were presented in a block design with three alternating blocks of 2 min listening task and 2 min driving without any task. In addition to the secondary tasks, a tertiary task was continuously presented following a subsidiary task paradigm.

Secondary auditory workload task.

For the auditory listening task, six 2-min long story parts from the German audio narration "Russendisko" (Kaminer, 2000) were presented via loudspeaker within the vehicle. The selected monologues included no additional sound effects and they were modified beforehand such that the voice of the male narrator was superimposed with the voice of a female newscaster. The subject was instructed to ignore the latter and to carefully listen to the story, which was verified by asking three multiple-choice test questions (three answer alternatives, only one correct) about details of the semantic content of each story section at a turn point on the test course. The questions referred to a specific detail of the story and therefore allowed to control that the drivers were really listening to the task.

Secondary mental arithmetic task.

Subjects were asked to silently count backward from a large three-digit number in steps of 27 as fast and accurately as possible. At the beginning of the task, the number was announced by the investigator. Accuracy was tested by prompting the final number at the end of each block. The instructor set markers into the EEG to ensure that data sections during which subjects were questioned could be excluded from the analysis. For the presented numbers it was guaranteed that while counting down in steps of 27 no number could be encountered twice. The following numbers were used in the presented order: 889, 917, 783, 839, 867, and 787.

Tertiary auditory stimulus-response task.

In the tertiary task "left" and "right" German auditory commands were pseudo-randomly presented with a ratio of 50/50 chance of occurrence every 7.5 s. Stimuli were replayed from a predefined list that was similar for each subject. The stimuli had to be acknowledged by pressing the corresponding left or right button attached to each index finger as quickly and accurately as possible. Speed and accuracy of responses for the tertiary task were measured as performance measures for the task.

4.2.5 Experimental Procedure

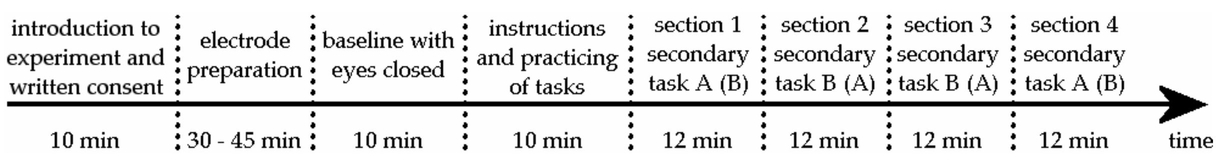

Figure 27. Schematic overview of the experimental procedure in Experiment 2. Each section refers to a drive from Esslingen towards Göppingen or the other way back. The order of tasks was counterbalanced across subjects so that tasks were alternated over the two different road sections. During each section three 2-min task blocks and three 2-min blocks without task were alternated. See text below for details.

Figure 27 displays a schematic overview of the experimental procedure. Subjects were selected on the basis of a screening questionnaire and assigned a date and time for their participation. Two testing time slots were available per day. Drivers were randomly assigned to either the morning (8 am - 12 pm) or the afternoon (2 pm - 5 pm) time slot. On arrival, there was a brief description of the experimental procedure and the vehicle that the subjects would be driving. The drivers were informed about potential risks and their responsibilities involved in the participation in this real-traffic experiment. It was pointed out that traffic violations were the drivers own responsibility and that the participation in the experiment would not take away any of the drivers' responsibilities. It was emphasized that no matter what secondary task would be presented, the drivers had to ensure by themselves that their first priority lied on safe driving in traffic. Next the EEG and ECG were prepared in a separate office room which took about 30 - 45 min. This also included a short demonstration of eye movement and muscle artifacts to improve subjects' compliance with the experimental procedure, e.g. drivers should not talk during the actual experiment. Every experiment started with a 10-min baseline EEG recording that took place inside the vehicle in the underground garage of the office building. During this baseline participants had to close their eyes and relax. For the recording of this baseline, the instructor left the participant alone inside the car with turned-off engine. After that the tasks were explained to the drivers and practiced until the participants declared that they felt comfortable with the tasks. An instructor was driving with the subjects at all times to operate the stimulus presentation and monitor the data recording. The experiment started as soon as the drivers had entered the highway in Esslingen and had integrated themselves into the ongoing traffic on the right lane. Throughout the entire experiment, the instructor logged the beginning and end of secondary tasks manually by setting labeled markers into the EEG data. A stop-watch with countdown alarm was used to ensure an accurate timing of the experimental blocks in the secondary tasks. Each driver drove two rounds on the test course resulting into four sections between the three turns. Ten participants drove in the morning and performed the listening task on Section 1 and 4 and the mental arithmetic task on Section 2 and 3. For the 8 subjects who were tested in the afternoon this order was reversed, i.e. they performed the mental arithmetic task on Road Section 1 and 4 and the listening task on Section 2 and 3. Both tasks followed a strict block design with six blocks (three blocks of high workload, three blocks of driving only) on each road section. The tasks were presented in blocks of 2 min with a 2-min pause in between. In some rare occasions the current traffic situation required to temporarily pause the stimulus presentation and continue after the road was clear again. Those events were manually logged in the EEG file and the separate data were merged offline for the analysis.

4.2.6 Experimental Design

Table 5. Overview of the experimental design of Experiment 2.

	Section 1 and 4 or Section 2 and 3		Section 2 and 3 or Section 1 and 4	
Driving on highway	X	X	X	X
Secondary loading tasks				
(1) Story listening	X			
(2) Mental arithmetic			X	
Tertiary subsidiary task	X	X	X	X

Note: All 18 subjects were tested in all four experimental sections in a complete repeated-measures design. Each secondary task consisted of three 2-min blocks. The tertiary task was continuously and simultaneously presented.

Table 5 displays the experimental design of Experiment 2. The sole main factor consisted of the mental workload manipulation which was induced by either the story listening task or the mental arithmetic task. Task blocks of high workload were compared to blocks during which the subjects drove without additional task. In addition a tertiary auditory stimulus-response task was continuously presented throughout the experiment and in parallel to the four workload manipulations. Every subject drove all four driving sections from Section 1 to Section 4 (see previous section 4.2.5 for more details on the experimental procedure).

The different behavioral, neurophysiological, and peripheral physiological measures that represented the dependent variables in Experiment 2 are listed in *Table 6*.

Table 6. Overview of the different dependent variables of Experiment 2.

Behavioral performance	
	Percentage of correct answers about story content
	Percentage of correct mental arithmetic results
	Tertiary task performance
	Reaction times
	Accuracy (d′)
EEG	
	Individual and task-specific alpha power amplitude
	P300 amplitude
	N1 amplitude
ECG	
	Heart rate (bpm)
	Heart rate variability
	Standard deviation of N-N intervalls (SDNN)
	Percentage of N-N intervals with at least 50 ms deviation from the preceding N-N interval
	Root mean square of successive N-N interval differences (RMSSD)

4.2.7 Data Analysis

4.2.7.1 Behavioral data analysis.

Mental arithmetic task performance.

After each 2-min task block the driver was prompted for the last result. The drivers' answers were compared with a list of possible multipliers and the responses were written down by the instructor in the car. For each driver the overall percentage of correct answers was calculated offline.

Auditory listening task performance.

The drivers' answers to the questions were logged and the percentages of correct answers were calculated for each driver. Due to a recording failure, the last three blocks' performance could not be evaluated for two drivers in this task.

Tertiary task performance.

All responses occurring between the start and the end of the secondary task were entered into the analysis. Correct answers were counted for reactions that occurred within a given time window of 200 - 3000 ms after stimulus onset. Impossibly fast answers, i.e. < 200 ms were excluded from the analysis. If two answers were given the first was evaluated and the second answer ignored. The answer to the commands could either have been a correct response, a false response, or a miss. In order to calculate d' the reactions to the "right" command were arbitrarily chosen to be counted as hits. In this way, correct answers to the "left" command were counted as correct rejections. The accuracy measure d' and the reactions times were similarly calculated to Experiment 1 (see section 3.2.7.1). One subject had to be excluded from the accuracy analysis of d’ due to failure of the left response button. The mean average for each driving section has been reported in the results part below.

4.2.7.2 Neurophysiological data analysis.

Overall EEG alpha band analysis.

Different from Experiment 1, the data was pre-processed in the Brainvision Analyzer software. Markers showing the start and end of each experimental condition were inspected for correctness and critical markers were corrected by inspection of the simultaneously recorded videos. After that, the data were high-pass filtered at 1 Hz, resampled to 125 Hz and then exported to the EEGLAB toolbox in MATLAB. Next, timing of the experimental blocks was reconstructed from the data and carefully checked again for plausibility by considering the experimental block durations. Foloowing this, the EEG data was segmented into the experimental blocks. One block fragment with less than 60 s duration was excluded from further analysis. Two blocks of another subject had missing data. The following data processing steps were similar to Experiment 1 encompassing analysis steps of re-referencing, artifact handling, data segmentation, power spectra computation, averaging over blocks, individual and task-specific alpha peak adjustment, and export to SPSS for statistical analysis (see section 3.2.7.2).

Analysis of event-related potentials.

For the analysis of event-related potentials, the preprocessed EEG data was loaded into MATLAB and hit responses were identified. Similarly to the spectral analysis, 11 electrodes (F3, F4, C3, C4, P3, P4, O1, O2, Fz, Cz, and Pz) situated in the middle of the scalp and the EOG electrode below the right eye were selected for further processing.

Since both stimuli consist of one syllable, no differences were visible between the averaged signal at all electrodes to “left” and “right” commands. As a consequence, the data for both stimulus types were combined. Following this, the EEG data was segmented into epochs of 1000 ms length around the stimulus event that preceded a hit. A pre-stimulus interval of 200 ms was chosen and the data were baseline-corrected by subtracting the mean amplitude of this pre-stimulus interval. During these preprocessing steps, the raw data was inspected and intervals were marked that contained artifacts. Data epochs that fell into these marked intervals were eliminated from the data. Next, the eleven electrodes of each trial were plotted together with the EOG channel and visually inspected for artifacts. Trials containing artifacts were rejected manually in this step. The mean reject rate over all subjects based on the visual inspection of the EEG trials was 12.38 % (range = 0.31 % - 26.85 %). The remaining epochs were averaged for each subject and then averaged over all subjects to obtain a grand average. Standard time windows for the N1 (80 - 170 ms) and P3 (300 - 600 ms) components were used to extract the mean amplitude data which were exported for the statistical analysis.

4.2.7.3 Peripheral physiological data analysis.

The analysis procedure for the two ECG channels was similar to Experiment 1 (see section 3.2.7.3). Generally, only the ECG electrode at the upper end of the sternum was used for the analysis. For 1 subject the second ECG channel had to be taken due to missing data in the first channel. After screening of the means over all subjects, 2 more subjects with extraordinary high values (M_{s08} = 73.69, M_{s13} = 50.43 compared to M_{all} = 19.46, SD_{all} = 25.53) that exceeded the predefined data limits (see section 3.2.7.3) were excluded from the PNN50 analysis. Mean values of the four measurements for each experimental block were exported for the statistical analysis.

4.3 Results

All statistical tests reported in this section assumed an α-level of .05. If ANOVAs showed a violated sphericity the Greenhouse-Geisser corrected p-values (pGG) were reported.

4.3.1 Behavioral Data

The behavioral data encompassed the performance for the secondary loading tasks, i.e. the results from the mental arithmetic task and the answers to questions about the story content. Moreover, the results from the responses to the stimuli in the subsidiary tertiary command task are reported.

4.3.1.1 Mental arithmetic and story recall performance.

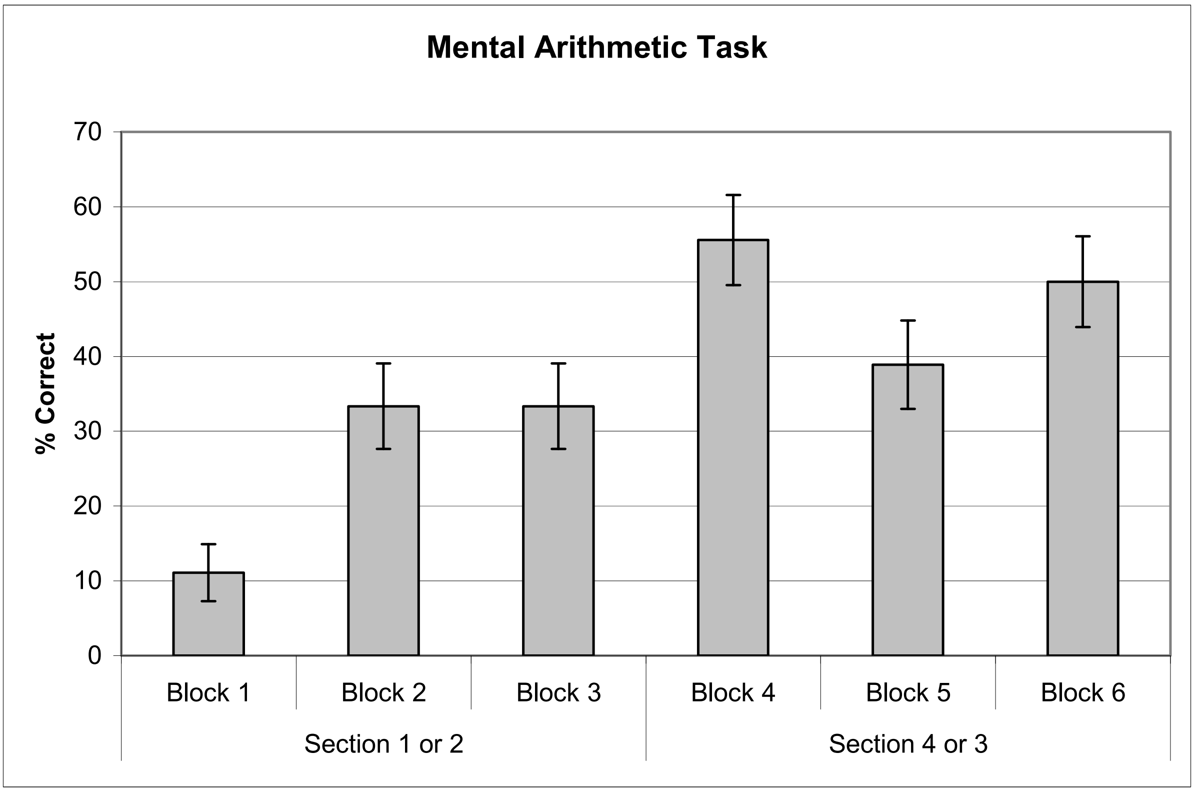

Figure 28. Average percentage (n = 18) of correct results in the mental arithmetic task. Please note that blocks 1 - 3 refer to a different driving section than blocks 4 - 6. Error bars represent the standard errors.

As can be seen in *Figure 28,* calculating downwards in steps of 27 was fairly difficult while driving. Over all eighteen subjects an average performance of 37 % correct answers (range = 0 - 83 %) was achieved. In addition, a learning effect can be observed over blocks ($M_{Block\ 1-3}$ = 26 %, $SE_{Block\ 1-3}$ = 7 % vs. $M_{Block\ 4-6}$ = 48 %, $SE_{Block\ 4-6}$ = 4 %).

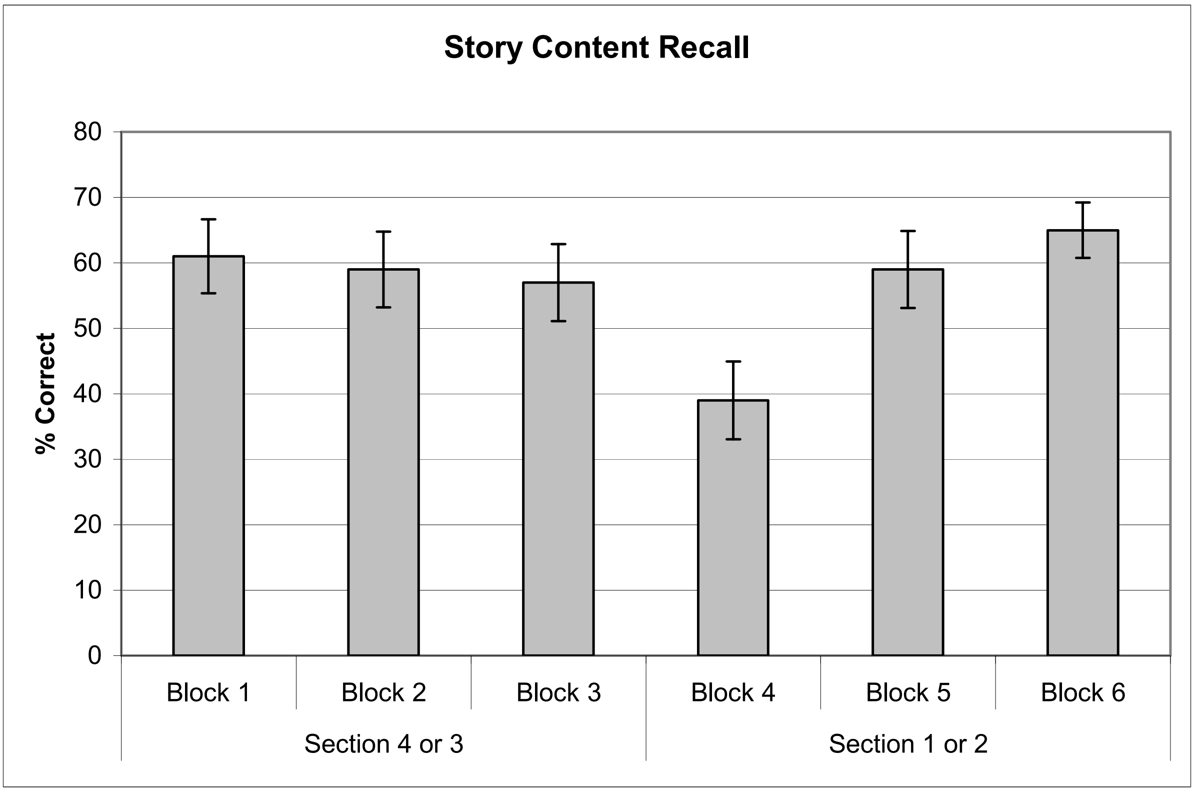

Figure 29. Average percentage of correct anwers ($n_{Block\ 1-3} = 18$, $n_{Block\ 4-6} = 16$) to the three multiple-choice content questions. Please note that Blocks 1 - 3 refer to a different road section than Blocks 4 - 6. Error bars represent the standard errors.

Figure 29 shows the percentage of correct answers per triplet of questions in each block for the listening task. An overall mean of 61 % (range = 28 % – 100 %) correct answers was achieved by the drivers. The bad performance in the fourth block may stem from an unusually difficult first question in the triplet that only 17 % of the subjects answered correctly.

The performances in the secondary tasks show that drivers were really engaged into these tasks and splitted their attentional resources between the driving and the additional cognitive task. It can be assumed that the tasks were sufficiently difficult and that subjects were mentally strained by performing the secondary tasks. While exiting the highway and turning the car, subjects sometimes used the short break to give comments that were recorded by video, e.g. "Oh my god, this is indeed very straining" [German: "Oh Gott, des schlaucht ganz schoen"] (Subject 12, video time of day = 10:06:07 am).

4.3.1.2 Tertiary task performance.

The mean reaction time and mean d′ over all blocks and subjects for the responses to the tertiary stimulus-response tasks are shown in *Figure 30* and *Figure 31*.

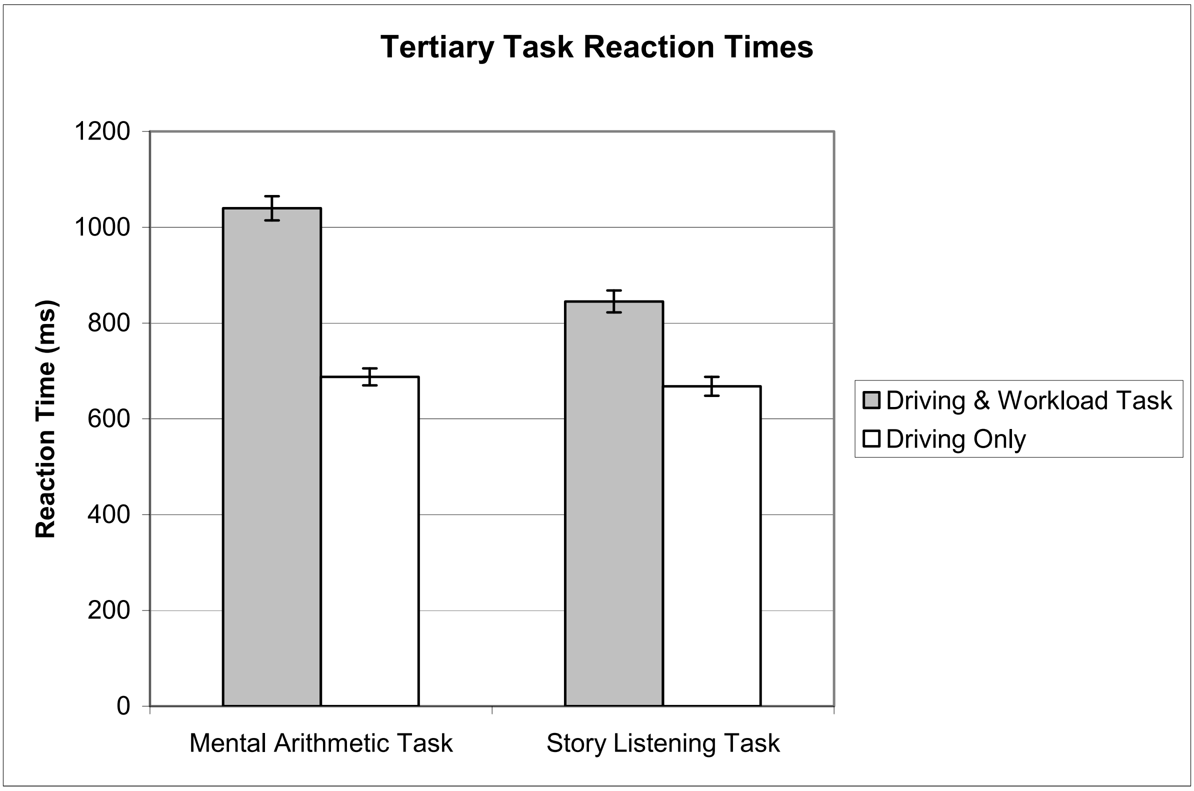

Figure 30. Mean reaction time values for the tertiary stimulus-response task averaged over all six blocks and all subjects (n = 18). Error bars represent the standard errors.

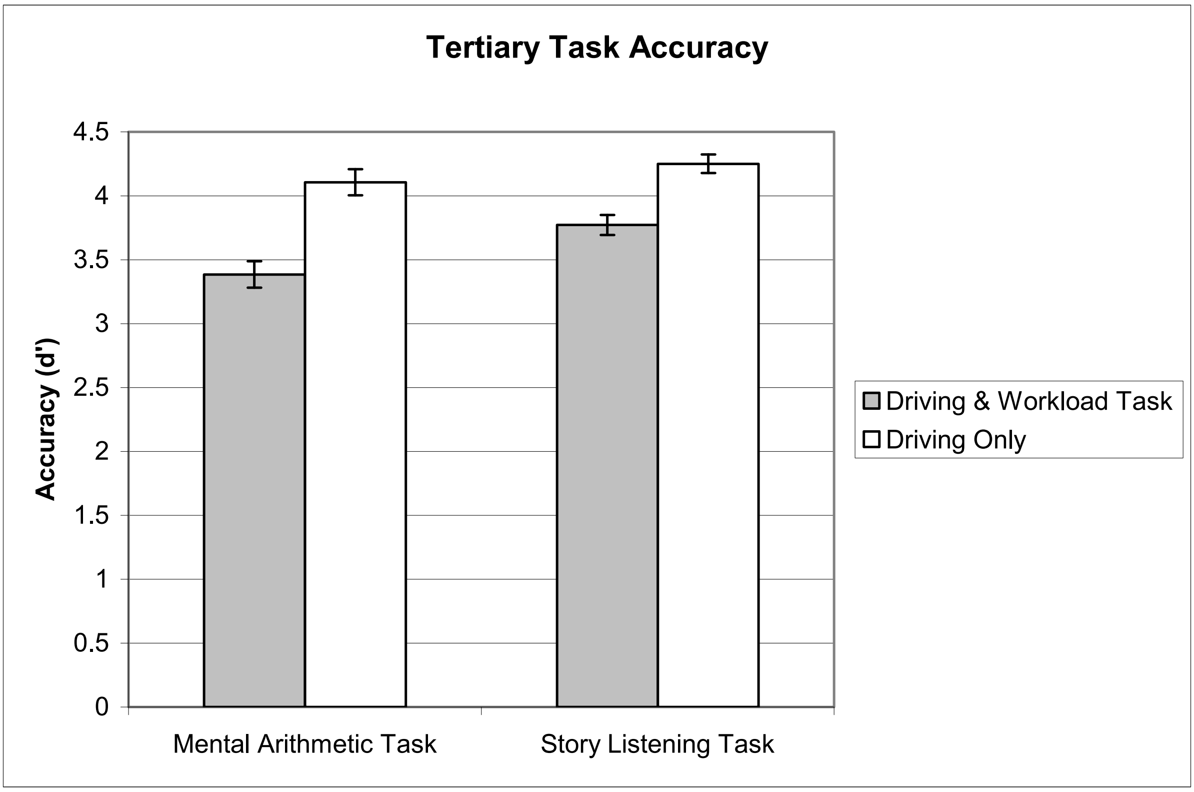

Figure 31. Mean d′ accuracy values for the tertiary stimulus-response task averaged over all six blocks and all subjects (n = 17). Error bars represent the standard errors.

Although the task was very easy as indicated by high d′ values, the evaluation of the performance measures reflected the influence of task difficulty onto the drivers. Reaction times show that subjects clearly responded later and made more false responses when performing a cognitively loading task. Moreover, although the effect of mental workload is evident in both secondary tasks, it is more pronounced for the mental arithmetic task than for the story listening task. This effect is more distinct for the reaction times than for d′. As expected there is apparently no difference between driving only conditions by looking at the mean values.

Reaction times and d′ were analyzed in two separate two-factorial repeated-measures ANOVAs with the main factors “task difficulty” (secondary loading task during driving vs. driving without additional task) and “type of task” (story listening task vs. mental arithmetic task). The analysis confirmed the observations for the reaction time means. The evaluation over the 18 subjects revealed a stable task difficulty main effect ($F(1, 17) = 135.286$, $p < .001$, $\eta^2 = .888$). A comparison between the two workload tasks showed that the mental arithmetic task produced significantly higher reaction times than the listening task ($F(1, 17) = 36.602$, $p < .001$, $\eta^2 = .683$). A significant interaction was found between the two main factors which most likely stemmed from a greater discrepancy between the two task load conditions in the mental arithmetic task than in the listening task ($F(1, 17) = 55.096$, $p < .001$, $\eta^2 = .764$). The means indicate longer reaction times if the participant had to count downwards in steps of 27 than when following the content of a story. The two baseline conditions between the cognitive task blocks showed only marginal differences.

17 subjects were analyzed in the ANOVA for the measure of detection accuracy (d'). Only the task difficulty factor was significant indicating that both loading tasks led to less accurate responses than when subjects drove without task ($F(1, 17) = 34.442$, $p < .001$, $\eta^2 = .683$). The type of task factor just failed to reach significance ($F(1, 16) = 4.188$, $p = .058$, $1 - \beta = .985$ for $\alpha = .05$, $n = 17$, $\eta^2 = .207$, $r = .546$), although the means in *Figure 31* show a clear tendency towards a less accurate performance under high mental workload imposed by the mental arithmetic task as opposed to the story listening task condition. The interaction between main factors was not significant ($F(1, 16) = 1.265$). As in the laboratory experiment, the behavioral performance outcome indicates the successful task manipulation of mental workload during the experiment.

4.3.2 EEG Alpha Power

Similarly to Experiment 1, two sets of statistical analyses over 11 EEG scalp electrodes were run (see section 3.3.3).

4.3.2.1 Individual alpha peak adjustment.

This section describes the results of the alpha peak frequency range and the alpha peak power amplitude evaluation. Based on the results from Experiment 1, it has been assumed that the different characteristics of the alpha effects would require an individual and task-specific alpha peak frequency adjustment.

Individual alpha peak frequency range.

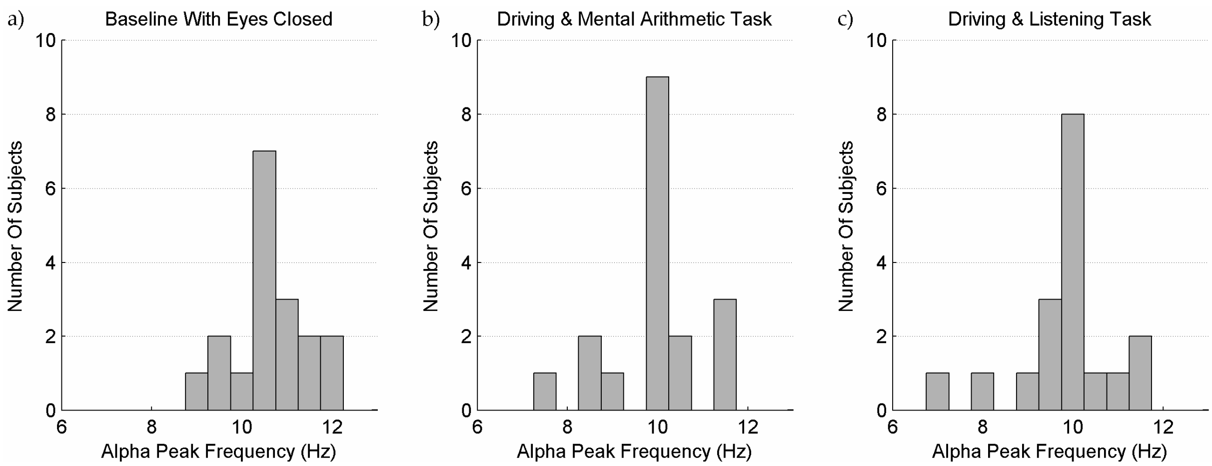

Figure 32. Distribution of individual alpha peak frequencies (n = 18) for the three different experimental conditions in Experiment 2.

Figure 32 shows the distribution over 18 subjects of the individual alpha peak frequencies in the different experimental conditions. As in Experiment 1, peak frequencies were determined by comparing all 11 electrodes and selecting the electrode with maximal power in the alpha frequency range. The distribution of alpha peak frequencies for the baseline with eyes closed in *Figure 32a* shows that most of the subjects showed a peak within a higher frequency range between 9 - 12 Hz, while peak frequencies in the two workload conditions during driving in

Figure 32b and *c* are more evenly distributed over the whole alpha frequency range. When looking at the mean alpha peak frequencies and their standard errors over subjects this observation finds additional support ($M_{baseline}$ = 10.63 Hz, $SE_{baseline}$ = 0.19; $M_{arithmetic}$ = 9.95 Hz, $SE_{arithmetic}$ = 0.25; $M_{listening}$ = 9.84 Hz, $SE_{listening}$ = 0.26). To test these differences for statistical significance, the individual alpha peak frequency values were entered into an univariate repeated-measures ANOVA. The "task condition" factor (baseline with eyes closed vs. driving and mental arithmetic task vs. driving and story listening task) revealed a statistically significant difference ($F(2, 34) = 9.881$, $p < .001$, $\eta^2 = .368$). Helmert contrasts showed that workload-specific peak frequencies did not differ ($F < 1$, $1 - \beta = .639$ for $\alpha = .05$, $n = 18$, $\eta^2 = .031$, $r = .821$), but that the significant effect actually stemmed from the baseline condition compared to the two workload-specific peak frequencies ($F(1, 27) = 14.133$, $p < .01$, $\eta^2 = .454$).

Individual alpha peak power amplitudes.

Figure 33 shows the individual mean alpha power peak amplitudes ($n = 18$) for the three conditions averaged over left and right electrodes. Power amplitudes indicate clear differences across task conditions. As in Experiment 1, the mean alpha amplitude for the baseline is much higher than the amplitude in the two demanding task conditions. Highest power values were recorded at parietal and occipital electrodes during the baseline condition. The two tasks show highest alpha power amplitudes over central and parietal electrodes. The story listening task elicited higher alpha power as opposed to the mental arithmetic task with the highest differences between tasks over central and frontal electrodes.

The highest amplitudes were obtained during the baseline, medium amplitudes were elicited during the story listening task, and lowest amplitudes were observed when subject performed the mental arithmetic task.

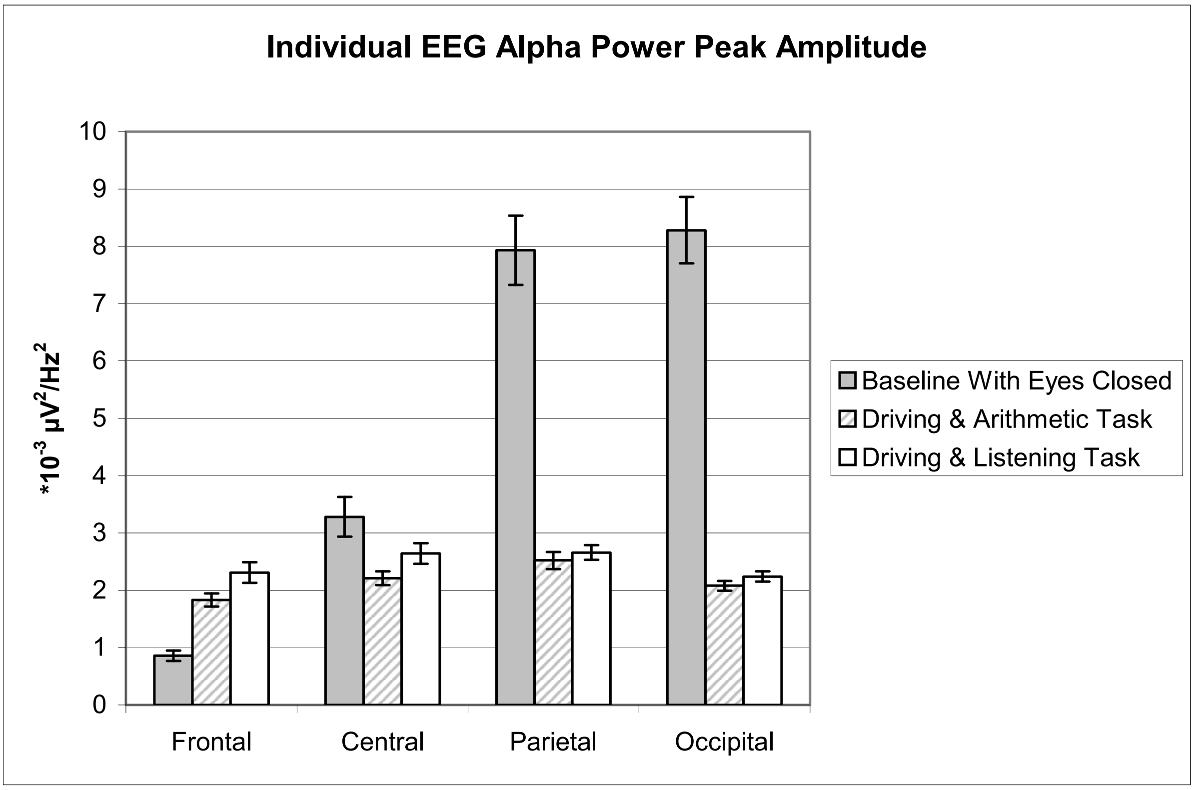

Figure 33. Mean individual alpha power ($n = 18$) recorded during the baseline with eyes closed or during driving and simultaneously performing a mental arithmetic task or a story listening task. Error bars represent the standard errors.

A repeated-measures ANOVA with the three main factors "electrode location" (frontal vs. central vs. parietal vs. occipital) and "type of task" (baseline with eyes closed vs. driving and mental arithmetic task vs. driving and story listening task) and "hemisphere" (left vs. right electrode) of the alpha power peak amplitude values revealed significant main effects for the task type factor ($F(1.063, 15.947) = 9.82$, $pGG < .01$, $\eta^2 = .396$), the location factor ($F(2.005, 30.079) = 23.209$, $pGG < .001$, $\eta^2 = .607$), and hemisphere factor ($F(1, 15) = 12.505$, $p < .01$, $\eta^2 = .455$). Helmert comparisons of the different levels for the type of task factor confirmed the observation that the resting baseline elicited higher power values than the two workload conditions ($F(1, 15) = 9.981$, $pGG < .01$, $\eta^2 = .4$) and that the story listening task elicited higher alpha power than the mental arithmetic task ($F(1, 15) = 4.747$, $pGG < .05$, $\eta^2 = .24$). A significant linear contrast for the electrode location factor reflected the strong occipital distribution of alpha power recorded during the baseline with eyes closed ($F(1, 15) = 41.681$, $pGG < .001$, $\eta^2 = .735$). A significant interaction between the type of task factor and electrode location factor ($F(1.588, 23.822) = 30.322$, $pGG < .001$, $\eta^2 = .669$) supported the topographical difference between the baseline alpha over parietal-occipital electrodes and the workload alpha over frontal-central electrode sites. A further investigation of the significant hemisphere main factor, i.e. the mean comparison between left and right electrodes, revealed that alpha power was stronger over the right hemisphere ($M = .0030$, $SE = .0003$) than over the left ($M = .0033$, $SE = 0.0004$). The source of this lateralization effect lay over parietal electrodes in the baseline condition with eyes closed ($M_{left} = 0.0067$, $SE_{left} = 0.0012$; $M_{right} = 0.0083$, $SE_{right} = .0014$; interaction between hemisphere, type of task, and electrode location: $F(2.654, 39.805) = 4.08$, $p < .016$, $\eta^2 = .214$).

The statistical analysis for the three midline electrodes just failed to reach significance for the type of task main factor ($F(1.1, 17.605) = 3.595$, $pGG = .071$) which was probably due to the required Greenhouse-Geisser correction for violated sphericity. The effect was clearly localized to electrode PZ in the baseline condition ($M_{frontal} = 0.0009$, $SE_{frontal} = 0.0002$; $M_{central} = 0.0027$, $SE_{central} = 0.0005$; $M_{parietal} = 0.0079$, $SE_{parietal} = 0.0014$). Consequently, the main effect of electrode location was significant ($F(1.053, 16.845) = 27.887$, $pGG < .001$, $\eta^2 = .635$).

It can be concluded that the different experimental conditions, the one being a resting baseline condition with eyes closed and the other being a driving condition with or without secondary tasks imposing high mental workload are fundamentally different. In this way, it has not been unexpected to see qualitatively different alpha power peaks in the power spectra of the individual drivers between the various conditions. Based on these observations, an individual alpha power adjustment is inevitable for the spectral EEG analysis.

4.3.2.2 Individual EEG alpha power effects of mental workload during driving.

The means of the individually adjusted alpha frequency power averaged over 18 subjects for the workload-relevant conditions and the distribution of the effects over the scalp can be seen in *Figure 34*. The listening task showed a very evenly distributed effect over all electrodes which was strongest at central and parietal electrodes. The mental arithmetic task revealed an incoherent pattern of results, i.e. an increase over parietal electrodes and a decrease over frontal electrodes. The listening task showed an overall higher alpha power level than the mental arithmetic task. By looking at the means, no clear differences could be observed between the two baseline driving conditions without additional task.

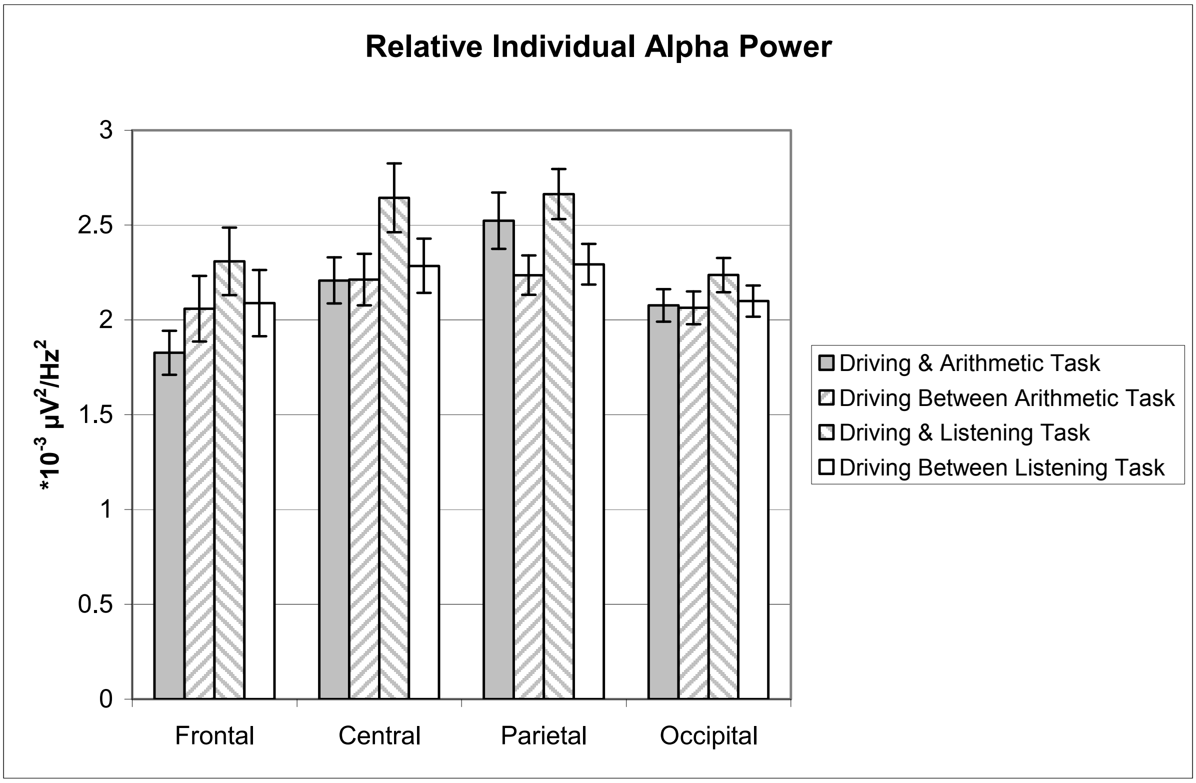

Figure 34. Mean individual EEG alpha power over subjects (n = 18) observed under different levels of mental workload (driving with or without secondary loading task) and for the two different types of cognitive task (mental arithmetic task vs. story listening task). Error bars represent the standard errors.

Two separate ANOVAS were run for each type of task (mental arithmetic task vs. auditory monitoring task). Each ANOVA included the two factors “electrode location” (frontal vs. central vs. parietal vs. occipital) and “task difficulty” (driving vs. driving with secondary task). The listening task revealed a statistically significant effect for the task difficulty main factor ($F(1, 17) = 10.323$, $p < .01$, $\eta^2 = .378$) supporting the observation that the individual alpha band power increased under high task load. The electrode location factor ($F(1.986, 33.764) = 1.029$) did not reach significance which speaks for the fact that equal levels of alpha power were recorded over all electrodes.

The analysis of the mental arithmetic task did not reveal any significant main effects for the task difficulty factor ($F < 1$) or the electrode location factor ($F(2.031, 34.524) = 1.825$). In case of the task difficulty factor this is probably due to the opposite direction of the effect over frontal and parietal electrodes (see *Figure 34*). In sum, only the story listening task revealed a significant and broadly distributed effect of mental workload during driving.

4.3.3 State Dependent Changes in Event-Related Potentials

Subjects continuously performed a simple “left-right” command task in parallel to the secondary workload tasks. *Figure 35* displays the grand average ERP data to the stimuli of this tertiary task and it compares the electrophysiological responses during driving with responses during driving and performing a mental arithmetic task in parallel. As can be seen, there was an increase in the N1 component around 150 ms over midline electrodes when subjects were under high workload. Moreover, a decrease in the P300 was observable at electrodes PZ, P3, and CZ.

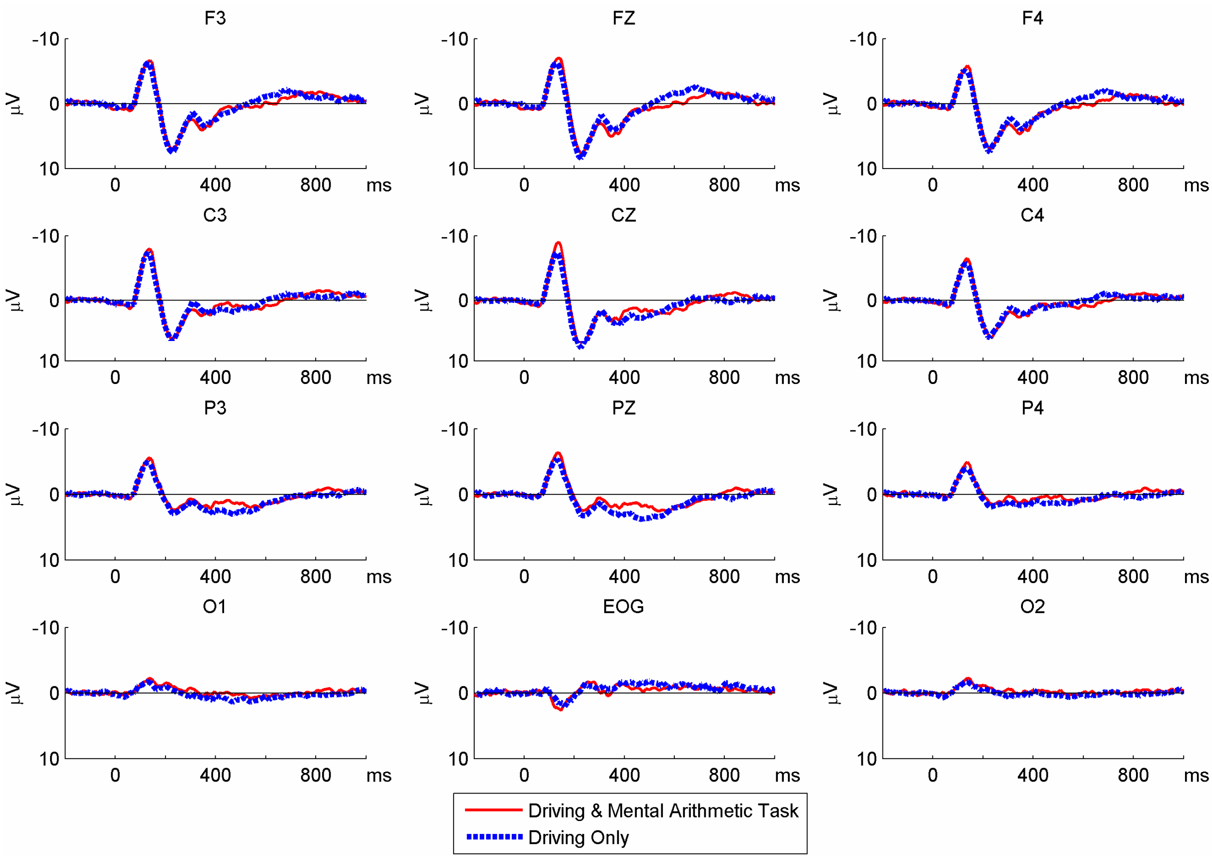

Figure 35. Grand average of the electrophysiological response to the stimuli of the tertiary command task (n = 18) while subjects were driving with or without performing a mental arithmetic task.

For the N1 component a repeated-measures ANOVA including the two main factors “electrode location” (FZ vs. CZ vs. PZ) and “type of task” (driving with secondary task vs. driving without additional task) was run. The statistics confirmed the N1 results reported above. Mean amplitude was increased over all three electrodes as supported by the two statistically significant main effects for the type of task factor ($F(1, 21.896) = 6.27$, $p < .05$, $\eta^2 = .269$) and electrode location factor ($F(1.288, 21.896) = 9.427$, pGG < .01, $\eta^2 = .357$). The effect was evenly distributed since no significant interaction between the type of task and electrode location factors was observed ($F < 1$).

The reduction of the P3 amplitude when subjects performed the mental arithmetic task was confirmed by the statistical analysis at electrode PZ. A repeated-measures ANOVA with the single main factor “type of task” (driving with mental arithmetic task vs. driving without additional task) was run and revealed a significant result ($F(1, 17) = 8.437$, $p < .05$, $\eta^2 = .332$).

Figure 36 shows the ERPs to the stimuli of the tertiary task while driving and listening to a story book in comparison to just driving. Most remarkable, when looking at the comparison between conditions in *Figure 36* a clear reduction of the N1 amplitude and around 100 ms a shift of about 50 ms length can be observed over frontal, central, and occipital electrodes in the driving and listening task condition as opposed to the driving only condition.

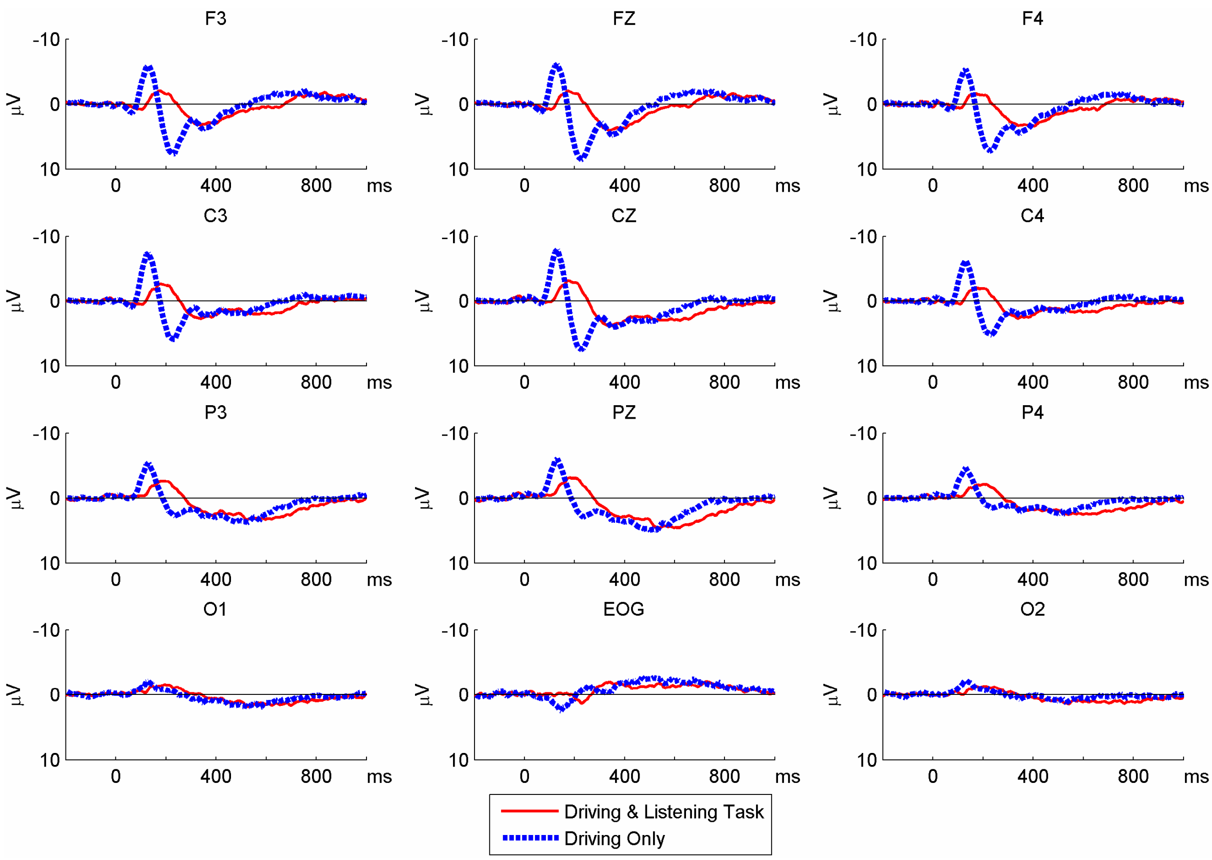

Figure 36. Grand average of the electrophysiological response to the stimuli of the tertiary command task (n = 18) while subjects were driving with or without performing a story listening task.

The extreme reduction and delay in the electrophysiological response in the N1 was confirmed by the significant results of a repeated-measures ANOVA with the two main factors "electrode location" (FZ vs. CZ vs. PZ) and "type of task" (driving and listening task vs. driving without additional task). Both the electrode location factor ($F(2, 34) = 13.714$, $p < .001$, $\eta^2 = .447$) and the task difficulty factor ($F(1, 17) = 101.283$, $p < .001$, $\eta^2 = .856$) were significant. The effect was stronger pronounced over central and frontal electrode sites than at the posterior electrode as indicated by a significant interaction between the type of task and the electrode location factor ($F(1.29, 21.923) = 7.808$, $pGG < .01$,$\eta^2 = .315$).

When comparing the waveforms of the two baseline driving recordings without additional task (see *Figure 35* and *Figure 36*) similar N1 - P2 complexes can be seen in the electrophysiological response to the command stimulus. A repeated-measures ANOVA with the factors "electrode location" (electrodes FZ vs. CZ vs. PZ) and "type of baseline driving" (baseline driving between arithmetic task performance vs. baseline driving between listening task performance) revealed no significant difference between conditions ($F < 1$) which speaks for the reliability of the measurement. The means indicated a more negatively pronounced N1 over central electrodes ($M = -4.78$, $SE = 0.41$) than frontal ($M = -3.82$, $SE = 0.38$) and parietal ($M = -3.8$, $SE = 0.31$) which was reflected by a significant electrode location factor in this analysis ($F(1.452, 24.69) = 8.214$, $pGG < .004$, $\eta^2 = .326$).

4.3.4 ECG Indices of Mental Workload

Mean values over all four blocks were calculated for the heart rate and each of the three heart rate variability dispersion measures. All mean variables were tested for normal distribution using the Kolmogorov-Smirnov test. Results showed normally distributed mean values for the variables heart rate and SDNN. Subsequently, these measures were evaluated using repeated-measures ANOVAs. The data of PNN50 and RMSSD did not fit a normal distribution and were therefore evaluated using the non-parametric Wilcoxon test for connected samples. None of the four Wilcoxon tests comparing each of the four conditions to each other revealed any significant results for both variables (driving - driving and mental arithmetic task: PNN50 $Z = -.22$; RMSSD $Z = -.157$; driving - driving and story listening task: PNN50 $Z = -.114$; RMSSD $Z = -1.817$). The heart rate and SDNN data were entered into separate repeated-measures ANOVAs with the main factors "task type" (mental arithmetic task vs. story listening task) and "task difficulty" (driving with workload task vs. driving without additional task).

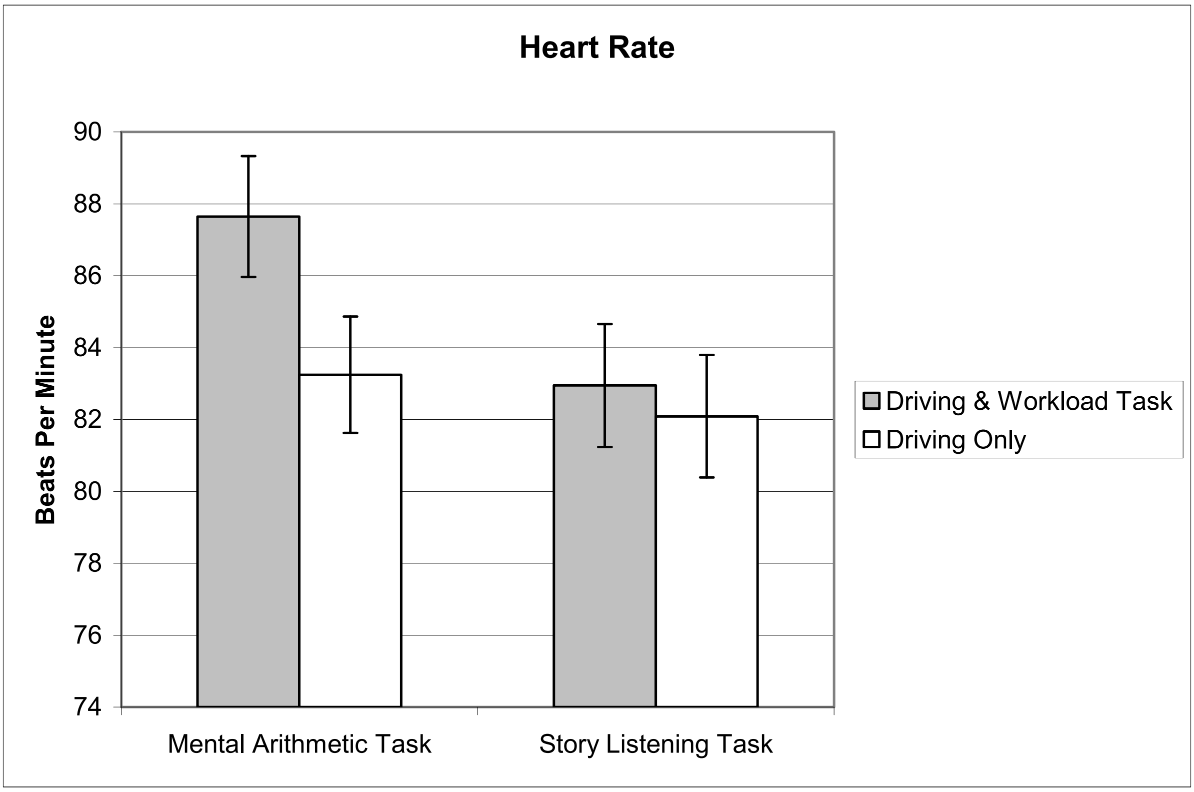

Figure 37. Mean heart rate ($n = 17$) for the experimental manipulation of mental workload (driving with or without secondary loading task) and the two different types of workload tasks (mental arithmetic task vs. story listening task). Error bars represent the standard errors.

As can be seen from the means displayed in *Figure 37*, the heart rate was clearly increased during performance of the mental arithmetic task while the other three experimental conditions did not significantly differ. This observation was statistically confirmed by a significant task difficulty factor ($F(1, 16) = 40.57$, $p < .001$, $\eta^2 = .717$) and the type of task factor ($F(1, 16) = 15.183$, $p < .01$, $\eta^2 = .487$). Likewise, the interaction between the two main factors was statistically significant ($F(1, 16) = 19.344$, $p < .001$, $\eta^2 = .547$) which indicated that the heart rate effect was greater for the mental arithmetic task compared to the story listening task.

Figure 38 shows that the mean SDNN values in the different experimental conditions. According to expectations, heart rate variability was significantly reduced when drivers drove with additional workload task ($F(1, 13) = 12.58$, $p < .01$, $\eta^2 = .492$) compared to driving without a task. No clear difference could be observed between the two types of secondary loading task ($F < 1$). The interaction between main factors was not significant ($F < 1$). While heart rate was only significantly increased during the mental arithmetic task, but not during the story listening task, heart rate variability measures reflected the influence of mental workload for both types of secondary loading task in the hypothesized direction.

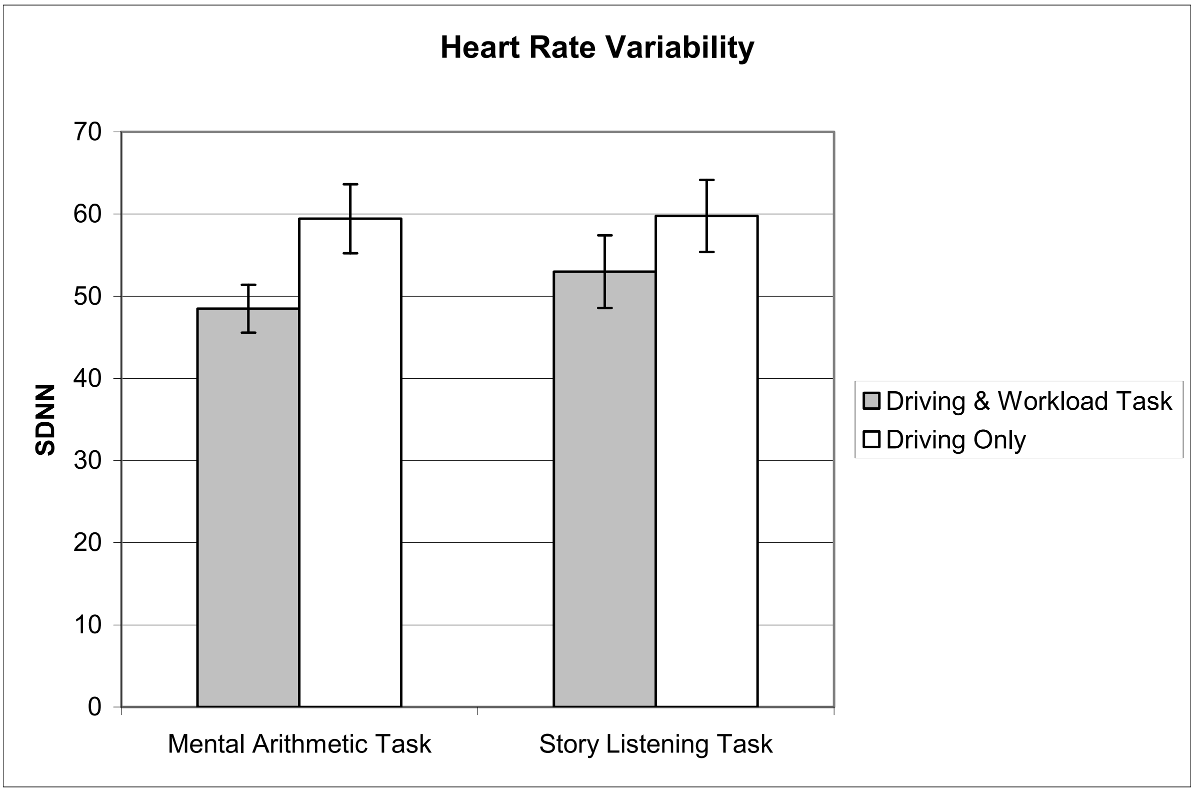

Figure 38. Mean standard deviation of N-N intervals (n = 14) in the ECG for the experimental manipulation of mental workload (driving with or without secondary loading task) and the two different types of workload tasks (mental arithmetic task vs. auditory monitoring task). Error bars represent the standard errors.

4.4 Discussion

The most important results from laboratory Experiment 1 is to demonstrate that mental workload induced by secondary tasks during a driving simulation leads to a significant increase in the EEG alpha frequency band. With Experiment 2 it has been tried to transfer these results to a real-traffic situation. Besides a modified version of the story listening task another cognitively demanding loading task was introduced to achieve a more detailed examination of drivers' mental workload. Results show that the alpha band is only sensitive to demands imposed by the story listening task. But at this point a newly introduced tertiary task provides valuable complementary information that exceeds the possibilities of an alpha frequency band analysis averaged over 2-min blocks. As it becomes evident from the overall pattern of results this field experiment combines several important advantages such as a guaranteed maximum subject involvement into the task situation, high face validity as well as the replication of previously observed effects under different situational conditions which speaks for the reliability and generalizability of the laboratory effects. The results will be evaluated one by one in the following discussion.

4.4.1 Interpretation of the Data and Implications

The data of the individual EEG alpha frequency band shows similarly to the laboratory experiment that intensive listening to a story leads to an increase in alpha power that is broadly distributed over all electrodes. Once more, this indicates the general inhibition of task-irrelevant brain areas during internalized attention. In contrast to this, the mental arithmetic task shows only a tendency towards an enhancement of individual alpha power over parietal electrodes, but this effect fails to reach statistical significance. This seems surprising in the light of previous laboratory experiments that very well demonstrated an influence of mental workload elicited by “internal attention” in a mental arithmetic task onto the power in the EEG alpha band (e.g. Ray & Cole, 1985b).

Nevertheless, there has also been a whole series of investigations that showed an alpha suppression for mental arithmetic tasks. This influence has even been demonstrated under dual task conditions (e.g. Ryu & Myung, 2005), but only in the laboratory. Up-to-date there is a lack of real-traffic driving studies. Despite contrary effects of single studies it can be argued that mental arithmetic tasks have been extensively studied in human cognitive performance studies. Following this, it seems quite unlikely that the alpha band is insensitive towards mental workload induced by such a task. But it may well be that drivers were not as strained by the task as it was initially planned.

This result may be related to the design of the task itself, which created a relatively open situation for the drivers. The subjects counted silently and were only prompted for their results after 2 min. Thus, the task continuous required high concentration of 2 min and at the same time it included only a low error tolerance for the task. For example, if a driver was distracted for just a few seconds by a traffic event and forgot the current calculation result, there was a high possibility that the participant was not able to get back to the task. On the one hand, an average of 37 % correct answers speaks for the high difficulty of the task, but on the other hand it probably also speaks for subjects’ failure in accurately performing the task for the full time span of 2 min.

Moreover, the inspection of the drivers’ answers to the contents of the presented stories indicates that despite the significant results there might have been a few deficits in the design of the tasks. While performers of the mental arithmetic task were asked for their results at the end of each block, the content questions were asked at the end of each driving section, i.e. when exiting the highway and making a turn. Worse recall performances for the fourth block may be related to the fact that drivers were simultaneously involved in a demanding driving maneuver, i.e. prompting them during exit and entry of the highway was doubtless not ideal. It might have been better to include a short stop at the turn points.

During the analysis of the alpha frequency band an individual and task-specific adjustment of the peak frequency has been performed. In contrast to the laboratory experiment, not only a difference in alpha amplitude, but also during the comparison of the peak frequency in the alpha band is observed. For the resting baseline, the peak frequencies lie in a higher range and are less broadly dispersed than in the two driving conditions under high workload. Moreover, the resting baseline shows considerably higher amplitude values which are strongest over parietal and occipital electrode sites. For the story listening task there is an overall higher alpha power level than for the mental arithmetic task. The two workload conditions do not differ in their mean peak frequency which speaks for a driving-specific shift in the power spectrum. From this initial analysis described above, it can be concluded that there is no straightforward relationship between the baseline alpha peaks and the alpha peaks elicited during driving under high workload. Therefore, it is not feasible to perform an individual alpha power adjustment based on baseline data recorded during a relaxed state. Another argument concerns the fact that it is clearly more difficult to obtain a clean measurement of the EEG during driving compared to the laboratory and it is shown that an adjustment procedure is indispensable. The existing effects would have probably been eliminated by a standard frequency band analysis with subsequent averaging over subjects due to the high amount of noise in the data.

When looking at the event-related electrophysiological potentials, the results draw a different picture for the two workload-inducing tasks which may also be related to task-specific stimulus modalities and different qualities of interference.

Both the content-monitoring task and the command task address the auditory sensory channel; hence based on existent multiple resource models (Wickens, 1984) interference between the secondary and tertiary tasks is very likely to occur. This interference is reflected by the ERP results in this condition: responding to commands and simultaneously listening to a narrative lead to an increased latency of about 50 ms of the N1 and a reduction in the component's amplitude compared to situations in which subjects heard the commands without the story. The decreased amplitude and the heightened latency during the workload condition support the idea of delayed information processing due to the difficult differentiation of stimuli and it seems to confirm existent laboratory results on divided attention (e.g. Kramer et al., 1988; Parasuraman, 1990; Wilson et al., 1994).

The ERPs during the mental arithmetic task show a different pattern of results. Although the arithmetic task and the command task do not occupy the same sensory bottleneck, both tasks interfere at the central processing stage. In fact there is an increase in the N1 amplitude which is normally regarded as a consequence of heightened attention. This seemingly initial antagonism to the ERP results of the other workload task may be easier to understand if considering the different task situation for the drivers. During the mental arithmetic task the prevailing silence inside the vehicle was only interrupted by the regular presentation of the left-right commands at intervals of 7.5 s. Since drivers fulfilled a self-paced task, it can be assumed that they controlled the tasks in a way that they could optimally respond to the recurrent commands without forgetting their current calculation progress. Otherwise the performance of these two tasks in parallel would surely not have been possible. Consequently, the drivers must have waited for the commands' appearance which explains the enhanced attention reflected by an augmented N1 amplitude.

A second result is shown by the predicted amplitude attenuation of the P3 during the mental arithmetic task. This result replicates widely reported findings stating that this component is sensitive towards perceptive-cognitive demands of a cognitively demanding task (see e.g. O'Donnell & Eggemeier, 1986; Tsang & Wilson, 1997). However, although the N1 effect seems to provide a definite result, the P3 outcome has to be interpreted with some caution. First of all, no classic oddball paradigm was employed, i.e. there were no rare stimuli among a series of frequently presented stimuli. For this reason, it was already assumed that effects would only appear moderately salient as opposed to an ERP elicited by a rare tone. It speaks for the high workload sensitivity of the measure that it nevertheless showed a statistically significant effect. Second, the objection is legitimate that due to the variation in the earlier component one has to be cautious in the interpretation of any following component. The difference in N1 amplitude can exert its influence onto the whole waveform so that there may be variations in later components that are not linked to the experimental workload manipulation.

Nevertheless, the question remains: How does this ERP result for the arithmetic task relate to the null results reported for the power of the alpha band in the same task? Unlike the frequency band analysis only the ERP demonstrate the consequences of mental workload for both tasks. At this point, one may want to conclude that the alpha band does not represent a suitable measure for high mental workload in a mental arithmetic task during driving. But this seems difficult to believe in the light of the other effects of the other secondary loading task onto the alpha band that are consistent with the hypothesis. There can only be speculations about the potential causes.

A possible explanation could be that the mental arithmetic task was too openly designed and that subjects probably didn't concentrate on the calculation in the predicted manner. However, this explanation does not necessarily exclude the possibility that participants were highly stressed, e.g. due to their attempts to remember a previous intermediate result in order to be able to continue with the task. Similarly to the discussion of the laboratory results, it is possible that by averaging over 2-min sections the smearing of the effect was too strong. Therefore, it can not be concluded that the EEG is generally insensitive. The stimulus-triggered ERPs are simply more sensitive and provide valuable additional information. The ERPs especially point out the fact that resources have different qualities for the two secondary loading tasks. While there are classic attention and workload effects for the mental arithmetic task, there is a comprehensive latency shift that seems to be completely based on different bottlenecks of cognitive processing. Nevertheless, alpha power should provide a more direct measurement of the driver's cognitive state while the ERP response is only indirectly related to the workload task, i.e. it is based on the assumption that the amount of cognitive spare resources can be inferred from the tertiary task.

Further evidence is provided by the behavioral measures of the tertiary task. In the present design, the tertiary task corresponds to the additional task in a subsidiary task design including a manipulation of task demands by the loading tasks. The pattern of results is according to the hypothesis. Results indicate that the spare capacity is reduced during the performance of the secondary task. Both tasks lead to worse detection accuracy and slower reaction times for the left-right commands. The reaction times in the mental arithmetic task condition even indicate a greater discrepancy between high workload and driving only conditions than in the comparison for the other task load condition.

Nevertheless, the very pronounced effects for the tertiary task indicate the actual degree of impairment for the drivers. Slow and inaccurate reactions as well as changes in the electrophysiological components that relate to attention processes as well as the encoding of new information fit into the overall concept of inattentional blindness (Strayer & Drews, 2007). Especially the case of concurrent auditory stimulation shows high face validity. Probably every driver can imagine having been in the situation to try to understand suddenly incoming traffic news from the radio while being in a conversation at the same time.

The last data source reported here concerns the ECG data. Although a respiratory control measure is missing, heart rate, and heart rate variability indicate to be sensitive measures. Results for the heart rate variability as measured by the SDNN are according to the hypotheses in this driving experiment, i.e. under influence of high mental workload, heart rate variability is decreased during the performance of both secondary loading tasks. Heart rate is clearly increased when the drivers performed the mental arithmetic task. Results are according to the general consensus in the literature (e.g. see Manzey, 1998). However, heart rate variability does not distinguish between different types of loading tasks despite the qualitative differences shown elsewhere. Therefore, it seems as if the sensitivity of the ECG measurement used in this work is rather unspecific. When comparing these results to the ECG results from the laboratory experiment the difficult question comes up why the results obtained from real-traffic are less ambiguous than the results obtained from the artificial environment. Maybe this observation is also owed to a higher task involvement in the field experiment.

4.4.2 General Discussion and Conclusions for Experiment 3

Finally some general points are brought up that concern Experiment 2 as a whole and that are of importance for the design of the final experiment of the present thesis. The driving experiment took place on a highly frequented highway which was surely not completely unproblematic. On the one hand, there was a myriad of influences by the other road users. For example, in one case the experiment had to be interrupted due to a traffic jam. On the other hand it became evident, that the choice of an efficient secondary task was strongly limited by the traffic situation. It is shown, that suitable tasks need to provide a very clear and fixed structure to really lead to an explicit outcome. Finally, it has also been for reasons of road safety that an area with restricted access and hence without traffic was selected for Experiment 3.

Another argument is concerned with the low number of tested subjects. It was not feasible to test a larger number of subjects due to the high investments for running such a driving experiment. That being said, it has to be admitted that additional participants would have enhanced the clarity of the results in Experiment 2. An a priori calculation of statistical power using G*Power based on a standard alpha and beta error level of .05, a correlation between repeated measures of $r = .5$ as well as $n = 18$ would have needed to assume an effect of $f > .45$, i.e. an effect that would have even been larger than what Cohen (1988) defined as “large” effect. Unfortunately, such effects could not necessarily be assumed offhand due to the difficult data situation in the field.

Some concerns relate to the stimulus material of the tertiary task that could have introduced two problems: (1) a possible interference with the driving task and (2) instead of using words, tones might have had the advantage to allow a cleaner comparability than words. With respect to these concerns, it can be argued that none of the drivers showed any interference that would have been critical for driving safety. During the experiment vehicle data were recorded on the CAN bus, which have been evaluated by collaborators in the scope of the overall project related to this experiment. Results showed no statistically significant effects of drivers’ steering movements as a function of workload manipulation by the secondary loading tasks (Dixon, K. R., personal communication, June 20, 2007). Responsible factors could have lain in the strongly controlled experimental driving situation. Moreover, all drivers reported to be comfortable with the task after a short practice period at the beginning.

The use of tones instead of words would have had two critical disadvantages. First, a direct mapping between the buttons at the left and right index fingers is easier to achieve with the semantic information that the stimuli contain. Second, tones are more salient from the constant narrative flow of the story teller and the news speaker. It remains an open question for future investigations whether this would have influenced the size of the effect. With respect to the perceptibility of warning signals within the vehicle this question can certainly be of interest.

In Experiment 3 the robustness of the effects found in the EEG alpha band has been evaluated under extreme driving conditions. By driving on dirt roads close-to-off-road conditions were created that pushed the EEG method to its limits due to the high amount of interspersions.

5 Experiment 3

5.1 Aim and Hypotheses

The final experiment in this thesis was run as part of a larger project that aimed at the investigation of driver and co-driver as a team. A larger number of subjects (30 teams) were tested. To report the whole experiment with all questions that it addressed would be beyond the scope of this work. For the present work, the data for the driver can deliver valuable extra insights on the electrophysiological response to mental workload induced by a secondary conversation-like task. Thus the following report focuses on the drivers while they performed only one cognitively demanding task, namely the same type of combined story listening and word detection task that was already used in Experiment 1.

However, as an expansion to the previous two experiments, drivers in Experiment 3 were tested in two sessions, one in the morning and one in the afternoon. Such a replication on the same day allows demonstrating the reliability of the effect. In addition to the baseline recording with eyes closed, this experiment included two extra baseline drives at the beginning and at the end of the experiment to enable a more thorough comparison between alpha power effects in a relaxed state and under task performance. Moreover, this final experiment took place without traffic, but in a sports utility vehicle on cross-country dirt roads. The purpose of this was to minimize environmental influences from sources such as the surrounding traffic to a minimum. Furthermore, the vehicle's governor was activated so that the drivers' speed could not exceed a predefined speed limit. Unfortunately, due to bad driving conditions during a heavy winter, driving on these snow-covered dirt roads also turned into a test for the robustness of the drivers' EEG signals. Besides the EEG, the ECG and the word detection performance data were evaluated to provide a comprehensive picture of the drivers' cognitive states.

This third experiment as it is reported here fulfilled only the single goal to provide additional evidence for the robustness and reliability of the demonstrated workload effects. It has thus been hypothesized that in both driving sessions, i.e. morning and afternoon session, drivers' electroencephalogram shows an enhancement of EEG alpha power during periods of mentally straining secondary task performance compared to just driving without additional task. Moreover, under periods of high workload the heart rate should be increased. Dispersion measures of heart rate variability (SDNN, PNN50, and RMSSD) are expected to show a decrease in variability under high workload. All results were expected to be replicable between the morning and afternoon sessions.

5.2 Methods

5.2.1 Subjects

30 employees of DaimlerChrysler were recruited as drivers through a general announcement. The subject population consisted of 22 males and 8 females ranging in age from 22 to 50 with an average age of 30 years. All participants were right-handed, except for one participant who was ambidextrous. For participation in the study, subjects were required to speak German as their first language and have no health conditions that might present a risk to safely completing the study or could potentially confound the results (e.g. hearing problems, psychotropic drug medication). An offline medical examination of the ECG data by a physician revealed no cardiac medical conditions for any of the subjects who agreed to take part in this assessment (Mücke, R., personal communication, July 24, 2007). All subjects possessed a valid driver's license and were regular drivers on a weekly basis at least. All drivers, except for one, had driving experience of at least six years. One driver had driven for four years. One subject took part in a prior pilot real-traffic driving study which involved similar tasks but a different driving route and different stimuli. The experiment was part of another study that required a simultaneous testing of driver and co-driver. Therefore, all except for the first driver who otherwise participated under exact same circumstances, were driving with a co-driver. After participation in the experiment, participants received a compensation for their efforts.

5.2.2 Experimental Setup

5.2.2.1 Test course.

The experiment took place at a disused German military base in Münsingen, Germany. The preparation of the recording as well as the resting baseline recording took place in a building next to the test course that was a 15-min drive away from the starting point of the first baseline drive. Eleven routes (mean length = 11.3 km, range = 9.4 km - 13.7 km) were selected that mostly included cross-country dirt roads, some passages through the forest and some sections on asphalt road. All roads were non-public, hence there was usually no other traffic. Each route contained five predefined zones with a mean length of 515 m (range = 119 m - 1206 m) during which the driver was acoustically triggered to slow down to walking speed. These zones were part of the co-drivers' task that was in the focus of a different research question outside the scope of this work. The set of routes included two baseline drives (morning route = 9.6 km, evening route = 5.9 km) on an asphalt road. A 2.9 km practice round helped the subjects to familiarize themselves with the tasks and the vehicle. This round was driven repeatedly until the subjects reported to be confident that they understood their tasks. It also included a 350 m long low speed zone in which the driver heard gun fire (see task description in section 5.2.4). The actual experiment reported here consisted of four 30 - 45 min rounds which were congruent in terms of driving difficulty and length. Each route was shown on a display to the driver while driving. Due to the strong winter, roads were partly covered with snow, although the routes were repeatedly cleared from the snow during the test campaign.

5.2.2.2 Experimental equipment.

The test-vehicle was a Mercedes Benz G500 (see *Figure 39*) modified to serve as a test-bed for various prototype technologies. The car was equipped with an automatic gear shift.

Figure 39. Mercedes-Benz G-class (photo copyright 2008 by the Daimler AG).

Figure 40. Custom-made dashboard of the test-vehicle in Experiment 3 including three displays, one for each car occupant and a middle display, that primarily showed a map with the vehicle's position and the current route. For the performance of the word detection task drivers had to press the green button at the dashboard located next to the steering wheel.

The custom-made dashboard (see *Figure 40*) included three 10′′ super wide SVGVA TFT displays (resolution: 1280 * 600 px): one in front of each front passenger and the third display located in the middle between the two. The driver's display showed a VAPS-generated standard driver instrument cluster (VAPS, Engenuity Technologies Inc., 2004). For each route a map was prepared on which the course was drawn in with a black line. GPS was recorded by a GPS mouse placed on the roof of the car. The GPS data was used to display a sector of about 2.1 km * 1.3 km of the map that was centered on the current position of the car. Throughout the experiment, the middle display showed a basic map with the current route and an arrow that indicated the present position and orientation of the car (see *Figure 41*). North was always displayed up on the map.

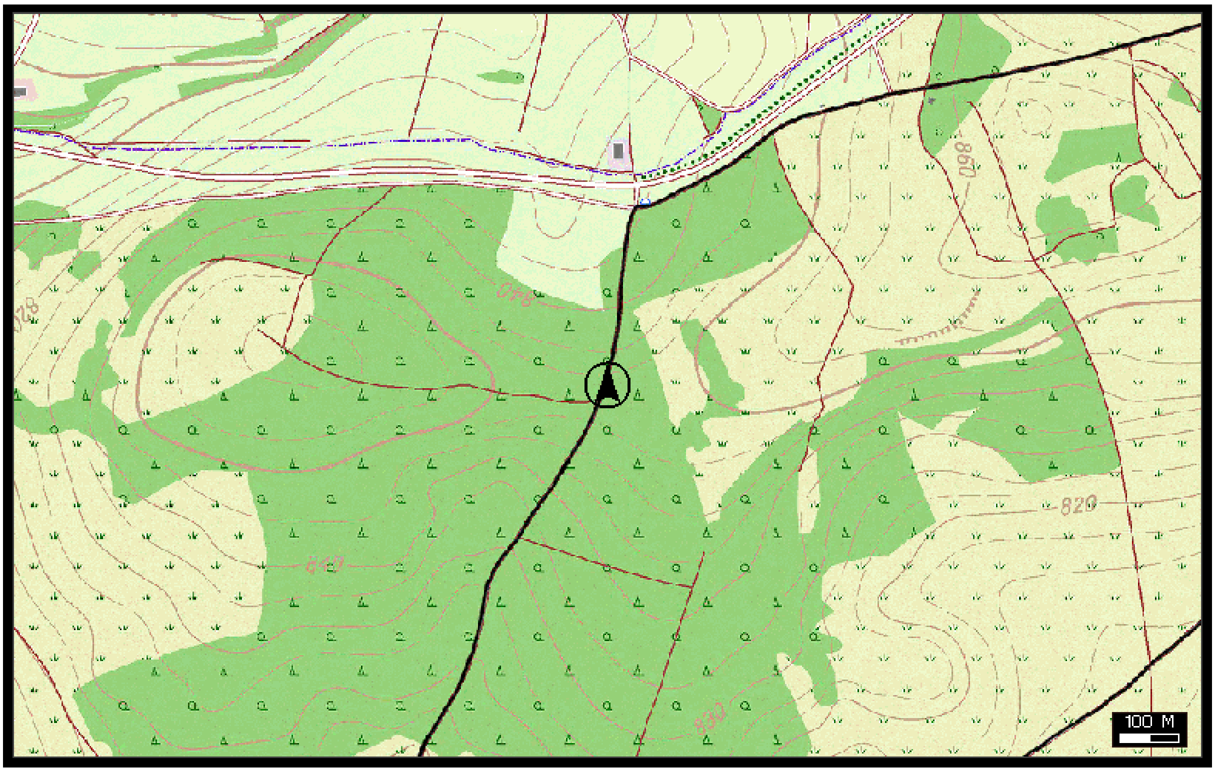

Figure 41. Exemplary map with route displayed on the middle display for the drivers. The arrow indicated the drivers' position and orientation on the test course as it was determined online by GPS.

Within the drivers' reach response buttons were installed to the right of the steering wheel on the dashboard. Due to project circumstances, steering of the car had an unusual momentum of about 190 - 200 Ncm (standard momentum < 160 Ncm), i.e. the driver needed to apply more force than usual when steering. Moreover, steering assistance was partly disabled, i.e. the steering wheel didn't move back like a standard assisted steering wheel. Therefore, drivers were warned at the beginning of the experiment about the unusual behavior of the steering when driving out of a curve. Moreover, due to additional 300 kg load on top of the roof, drivers were warned about potentially unusual driving dynamics of the vehicle, especially driving in curves on steep ascending slopes had to be avoided in order to not risk a fall over of the car. Portable BRAINAMP-amplifiers were installed on top of the Cardan tunnel in the back of the car together with the rest of the EEG equipment.

Figure 42. Trunk of the test vehicle with the six professionally built-in computers.

The vehicle had six computers installed in the back that were used for stimulus presentation and data acquisition. Although the machines were solidly and professionally built into the car, they were not completely shock resistant. Subjects were therefore informed to avoid deep holes in the road which would have endangered the sensitive computer hardware. The drivers were wearing headphones throughout the whole experiment. A safely installed camera inside the cabin which was synchronized with the EEG recorded a video that showed a view out of the front windshield.

5.2.3 Physiological Data Acquisition

The same 29 electrode EEG data recording setup as in Experiment 1 was used for each driver (see section 3.2.3) except for the fact that the driver's and co-driver's EEG were recorded together in one Brain Vision recorder workspace. Nevertheless, the two recordings were independent from each other and were cleanly separated prior to the analysis.

5.2.4 Task Description and Stimulus Material

Although roads were free from traffic, there were a few limitations compared to a naturalistic driving situation. Most of the time during the experiment, roads were covered with a layer of snow. Before driving on the test course, roads were sufficiently cleared from the snow to ensure that the drivers would not be handicapped. Although there was only rare traffic on the course, drivers were warned about eventual wild animals' crossing their path. Drivers were also asked to be careful to stay on the road since some of the terrain beside the roads was known to be contaminated with old ammunition. It was emphasized that no matter what task the drivers were encountering, they were instructed to put priority on safe driving first. While driving, the governor had to be set to 40 km/h at all times. Since the governor could be disabled by using the kick-down function, i.e. by fast acceleration, drivers were instructed how to reset the governor and to do so whenever necessary. The drivers' primary task was to follow the route shown on the middle display (see *Figure 41*). As the team traveled along the course, they encountered sections in which the co-drivers' task was to take pictures of target signs posted at the side of the road with a camera mounted on the top of the vehicle. Within these sections, the drivers had to keep the vehicle a stable platform and they were instructed to slow down and drive not faster than 10 km/h. The beginning of such a section was triggered manually by the instructor in the back of the car and the sound of gun fire salves was audible over the headphones throughout the whole target zone. Drivers and co-drivers were asked not to talk during the experiment and to avoid unnecessary movements not related to the driving to avoid artifacts in the EEG recording. While driving, a secondary task was introduced with the goal of deliberately inducing a specific type of cognitive load. The secondary task was presented in a block design with four alternating blocks of 3-min listening task and 3-min driving without the task per route. Subjects drove two routes in the morning and two routes in the afternoon.

The secondary loading task was a combined story listening and word detection task. For practical reasons it is also simply referred to as "listening task" in the following. For the listening task, the same 3-min long story parts from the German audio narration of the book "7 Jahre in Tibet" (Harrer, 1952) was used which had already been used in Experiment 1 (please refer to section 3.2.4 for details on the stimulus material). Four 3-min long narratives were auditorily presented on each of the four routes. A stimulus presentation program presented the stimuli via headphone to the drivers. Like in Experiment 1, the drivers' task was to listen to the content of the story and to respond to a certain keyword in the text. At the end of each route, the drivers had to answer one multiple choice question for each story section. The questions were the same as in Experiment 1, i.e. they referred to a specific detail of the story and therefore allowed to control that the drivers were really listening. For the detection of the keywords, drivers were asked to respond as fast and as accurately as possible by pushing a button on the dashboard with their right index finger. Drivers were instructed to keep both hands on the steering wheel and to only remove the hand for a response. As soon as the drivers started driving on a new route the present keyword was announced over the headphone followed by the beginning of the first story. In each round two 3-min story sections with a 3-min break in between were presented. Story blocks and the gun fire in target zones were presented independently from each other. Therefore, it happened that drivers heard a story section while they passed a target zone. However, the sound of the gun fire was reduced in noise level so that it was only audible in the background of a story.

5.2.5 Experimental procedure

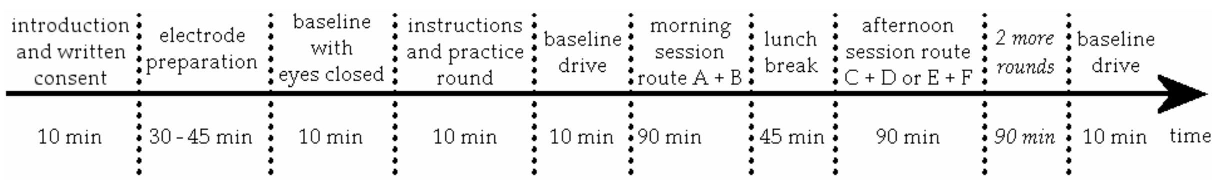

Figure 43. Schematic overview of the experimental procedure in Experiment 3. Two routes were driven in the morning and four routes were driven in the afternoon. Routes that were irrelevant for Experiment 3 are written in italic. Two sets of afternoon routes were alternately driven each day, i.e. half of the subjects drove route C and D directly after the lunch break and the other half drove route E and F. During each route four 3-min task blocks and four 3-min blocks without task were alternated.

The experimental protocol was officially approved by the US Western Institutional Review Board (WIRB Study No.: 1071995, WIRB PR. No.: 20051961). *Figure 43* displays a schematic overview of the experimental procedure for one driver. Prior to the experiment, every participant had received written general information about the tasks and goals of the study. The participants were informed that the experiment was designed to impose workload on the driver and co-driver via secondary tasks and that the experiment resembled a military setting, i.e. the experiment took place in a disused military facility and the stimulus presentation included the sound of gun fire. After setting an appointment, subjects were picked up in the morning at a predefined meeting point near Stuttgart and were driven about 80 km to the test course in Münsingen. After arrival at a barrack near the test course, subjects were once more informed about the experimental procedure and potential risks that stemmed from hazardous material lying off-road along the test course. Following this, both subjects signed their consent forms. After that, the physiological recording equipment was set up. Next, the volunteers were introduced to the vehicle (see description of vehicle in section 5.2.2.2). The experiment started with a 10-min EEG baseline recording during which the 2 subjects sat alone in the car with eyes closed and tried to relax. After this, the participants and the instructor drove to the starting point on the test track. This 15-min drive was also used to familiarize the drivers and co-drivers with the vehicle. Throughout the experiment, the instructor was sitting in the back of the vehicle at all times and operated the stimulus presentation and recording equipment. The experiment continued with a 9.6 km baseline drive on a plain asphalt road with rare traffic. After arrival at the starting point of the first practice round, subjects were instructed about the details of their tasks (see task descriptions in section 5.2.4). In the following, subjects drove two to three rounds on a 10-min practice round in order to get used to the driving and the additional secondary task. When subjects confirmed that they understood the task, the actual experiment started. In the morning session participants drove two rounds. After a 45-min lunch break at a barrack on the base, subjects continued with two more rounds on comparable rounds and with a similar stimulus presentation as in the morning. For the afternoon session, there were two sets of routes that were alternatively driven and two sets of stimulus material that were alternatively presented between subjects at different test days (13 subjects listened to the first set and 13 listened to the second set). After two more rounds which are not relevant to the data reported here, a last baseline drive was recorded from the subjects while they drove back on an asphalt road towards the barrack where the experiment had started in the morning.

5.2.6 Experimental Design

Table 7. Overview of the experimental design of Experiment 3.

	Morning session		Afternoon session	
	Route A and B		Route C and D or E and F	
Driving	X	X	X	X
Secondary loading task	X		X	

Note. The composition of each experimental condition is marked within each column respectively. All subjects were usually tested in all experimental conditions in both sessions and in a complete repeated-measures design. Each condition consisted of four 3-min blocks.

Table 7 displays the experimental design for Experiment 3. Every subject was tested in both driving sessions (for exceptions see section 5.2.1). The sole main factor consisted of the mental workload manipulation which was induced by the combined story listening and word detection task. Sections of high workload were compared to sections during which the subjects drove without additional task. Driving took place on different, but comparable routes.

The different behavioral, neurophysiological, and peripheral physiological measures that represented the dependent variables in Experiment 3 are listed in *Table 8*.

Table 8. Overview of the different dependent variables of Experiment 3.

Behavioral performance	
	Percentage of correct answers about story content
	Reaction times for word or tone detection
	Accuracy (d’) for word or tone detection
EEG	
	Individual and task-specific alpha power amplitude
ECG	
	Heart rate (bpm)
	Heart rate variability
	Standard deviation of N-N intervalls (SDNN)
	Percentage of N-N intervals with at least 50 ms deviation from the preceding N-N interval
	Root mean square of successive N-N interval differences (RMSSD)

5.2.7 Data Analysis

The following section describes the details of the behavioral, neurophysiological, and peripheral physiological data analysis. Due to failure of the technical equipment or no access to the test routes (heavy snow fall), three drivers could only participate in the baseline recording at the beginning. One subject could only drive the two routes in the morning, but not the routes in the afternoon. In the morning session 5 subjects drove only one route and for 1 subject the last story was not displayed. In the afternoon session, 2 subjects drove only one route and 1 subject had to finish after the first story block.

5.2.7.1 Secondary task performance analysis.

Both measures served as control measures for the subjects' pursuit of the secondary loading task. Please note that the detection performance contained no comparison for the experimental workload manipulation like in Experiment 1 (word detection vs. tone detection), but could only serve as a control measure to compare between test sessions.

Story comprehension.

At the end of each route drivers were asked one multiple choice question for each of the four stories that were presented. The percentages of correct answers over subjects were evaluated for each story block. Due to missing routes the subjects who completed all two morning routes (n = 16) and the subjects who completed all two afternoon routes (n = 12) were selected. Please note that such a procedure implicates that the analyzed drivers were not necessarily identical for the two sessions. Next, the average percentages of correct answers over all four questions for each route were calculated as well as the overall percentage of correct answers for all 12 subjects that completed all four routes.

Word detection performance analysis.

Due to unforeseeable events (e.g. driver missed a turn and had to drive back) in the experiment, the stimulus presentation had to be paused sometimes. This led to fractions of data that were merged prior to the analysis. Next and in order to obtain a measure for the key word detection performance, the 180-s story blocks of the word detection task were divided into consecutive 3-s windows starting from the beginning of a story block. As a result, stimulus time windows and non-stimulus time windows were obtained. Similar to the analysis in Experiment 1 and 2, d' was calculated based on theses time windows (see section 3.2.7.1). Time windows shorter than the length of 3 s were not included into the overall score. Responses later than 3 s after the stimulus were scored as a miss for the trial and counted as a false alarm for the following non-stimulus window. Overall 36 % of the stimulus intervals (53 of 149) had to be excluded from the analysis due to overlapping inter-stimulus-intervals of 3 s. In this case, the keywords appeared too quickly after another and it was not possible to assign a participant's response to the appropriate keyword. For the stimulus material of route A and B (see *Figure 43*), an average number of 12 stimulus (range = 4 - 16) and 29 non-stimulus time windows (range = 11 - 34) were analyzed per block. Over all blocks, responses to a mean of 86 key words (range = 48 - 99) and 210 non-key words (range = 110 – 239) per subject were entered into the analysis. For the stimulus material of route C, D, E, and F an average number of 12 stimulus (range = 6 - 17) and 29 non-stimulus time windows (range = 13 - 39) were analyzed per block. Over all blocks, responses to a mean of 87 key words (range = 51 - 97) and 219 non-key words (range = 112 – 236) per subject were evaluated. All reactions faster than 500 ms have been considered outliers and were excluded from the analysis. Due to the experimental setup it was very unlikely to react faster than this threshold.

Finally, the mean hit reaction times and mean d′ values were exported to SPSS for statistical analysis.

5.2.7.2 Neurophysiological data analysis.

Different from the analysis in the two previous experiments the EEG data was loaded into EEGLAB in MATLAB and the joint recording of both subjects was filtered (highpass: 0.5 Hz, lowpass: 56.25 Hz), down-sampled to 125 Hz, and re-referenced against the mean of each subjects' two mastoid electrodes (TP9 and TP10) respectively. Finally the data were each split up into two separate files, i.e. one for each person in the driving team. However, only the drivers' data were further processed for the present work. The analysis continued with the artifact rejection which was identical to the procedure reported for Experiment 1 (see section 3.2.7.2). As in Experiment 2, the timing of the experimental blocks was reconstructed from the data and carefully checked again for plausibility by looking at the experimental block durations. Next, based on the marker information of each experimental block, the EEG data was segmented into the experimental blocks. The following data processing steps were similar to Experiment 1 encompassing analysis steps of power spectra computation, averaging over blocks, individual and task-specific alpha peak adjustment, and the statistical analysis (for more details please refer to section 3.2.7.2).

5.2.7.3 Peripheral physiological data.

The analysis procedure of the ECG for this experiment was analogous to the analysis in the previously described experiments. Generally, only the ECG electrode at the upper end of the sternum was used for the analysis and the data was checked for outlier values that seemed physiologically implausible. (see section 3.2.3). One subject was manually excluded based on a final outlier inspection. The participant had an unusually low heart rate (M_{s32} = 50, SE_{s32} = 1.5) and an unusually high PNN50 (M_{s32} = 70.33, SD_{s32} = 2.06) in comparison to the other 26 subjects (heart rate: M_{all} = 86, SE_{all} = 2.21; PNN50: M_{all} = 7.85, SE_{all} = 1.44). Mean values of the four measurements for each experimental block were exported for the statistical analysis.

5.3 Results

All statistical tests reported in this section assumed an α-level of .05. If ANOVAs showed a violated sphericity the Greenhouse-Geisser corrected p-values (pGG) were reported.

5.3.1 Recall of Story Content and Detection Performance

Figure 44 displays the average percentage of correct answers to questions about the story for each of the four routes.

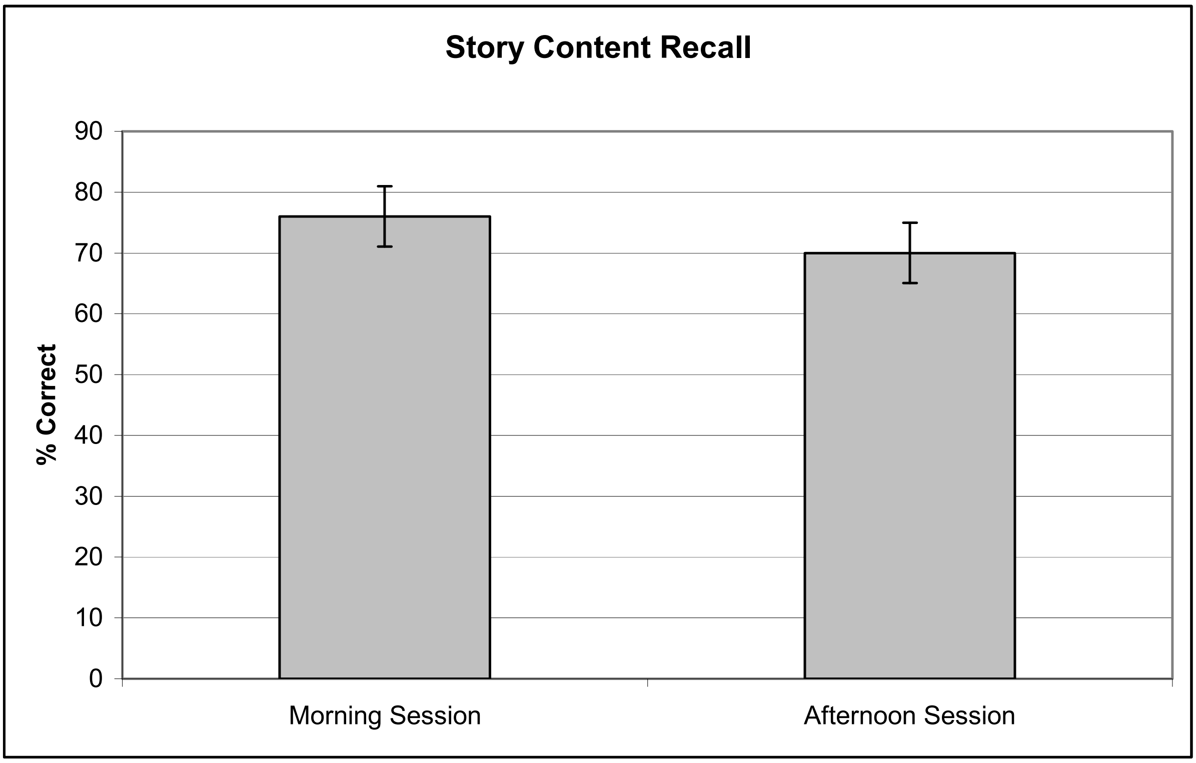

Figure 44. Average percentage of correct answers (n = 16) of four multiple-choice story content questions. Error bars represent the standard errors.

As can be seen, the drivers' answers were at a constantly acceptable level of above 60 % indicating that they really tried hard to follow the stories' contents while performing the word detection. For the drivers that completed all four routes a mean average of $M_{overall}$ = 72.9 % with a range of 43.75 % - 87.5 % correct answers between subjects was calculated. As in the previous two experiments, these data serve as a control by showing that all subjects listened carefully to the content of the story.

The mean reaction times and mean d' values over blocks for the auditory detection tasks are shown in *Figure 45* and *Figure 46*.

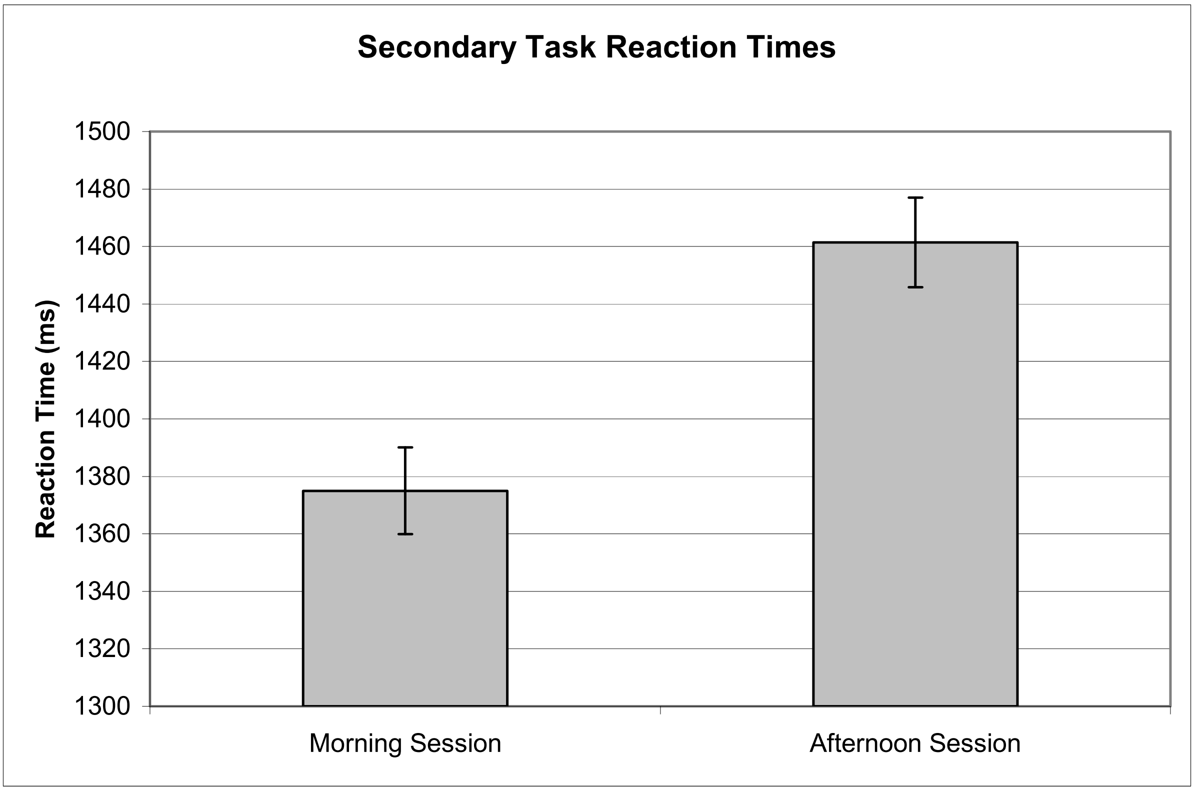

Figure 45. Mean reaction time values for the secondary word detection task averaged over all eight blocks and all subjects (n = 26). Error bars represent the standard errors.

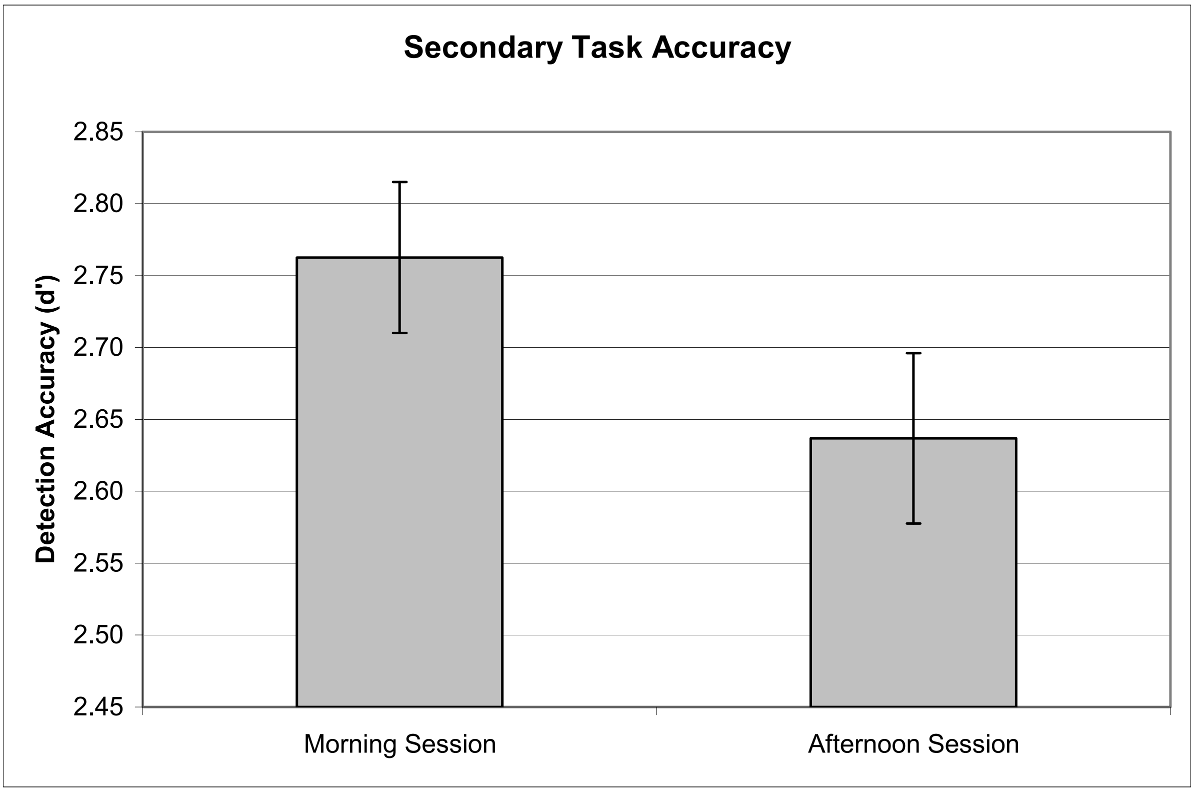

Figure 46. Mean d′ accuracy values for the secondary word detection task averaged over all eight blocks and all subjects (n = 26). Error bars represent the standard errors.

The detection performance in the afternoon session was slower and slightly less accurate than in the morning session. This observation was statistically confirmed by two-sided paired t-tests over 26 subjects between morning and afternoon measurements which revealed a significant difference in reaction times ($t(1, 25) = -5.319$, $p < .01$, $\eta^2 = .53$), but not for d′ ($t(1, 25) = 1.280$, $p = .212$; $1 - \beta = .329$ for $\alpha = .05$, $n = 26$, $dz = .243$, $r = .752$). Thus the detection performance data might speak for an influence based on fatigue in the afternoon.

5.3.2 EEG Alpha Power

Analogous to the previous experiments, two sets of statistical analyses were run for the 11 EEG scalp electrodes (see section 3.3.3).

5.3.2.1 Individual alpha peak adjustment.

Individual alpha peak frequency range.

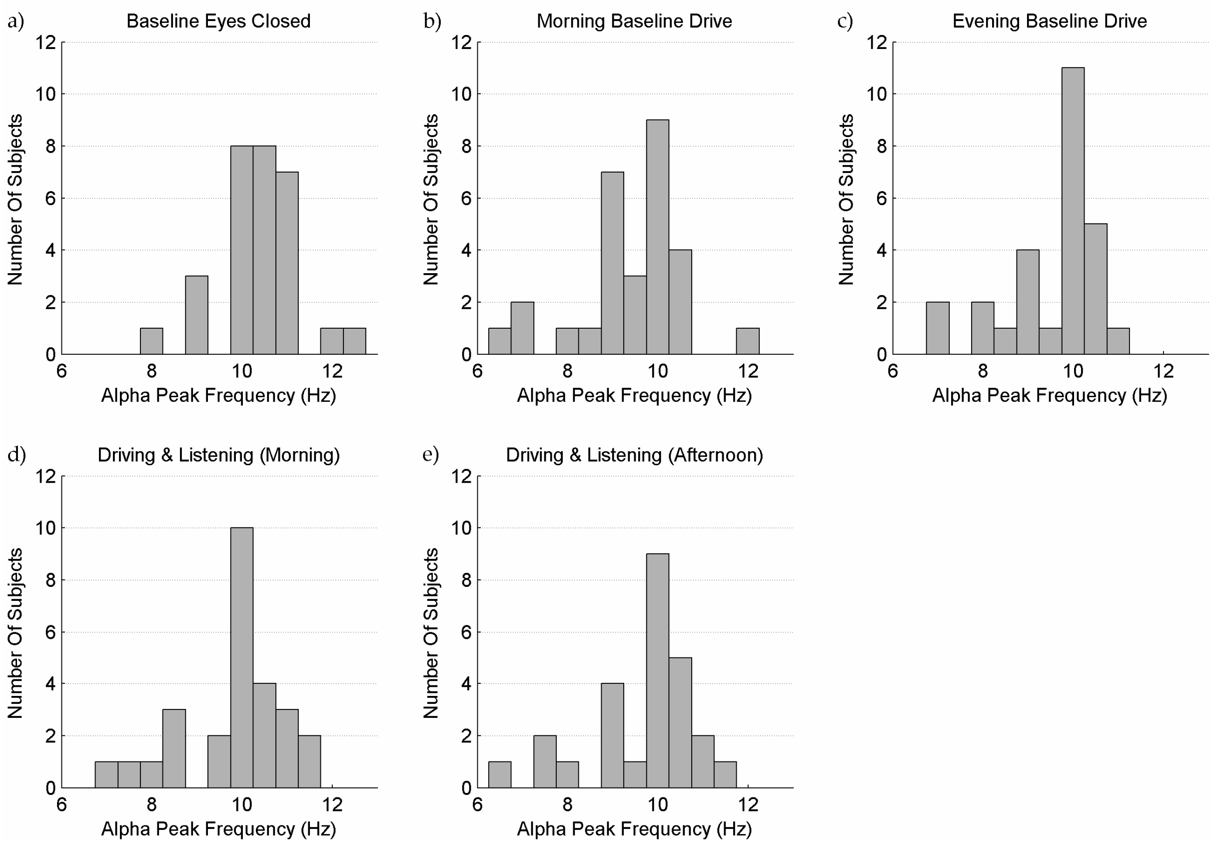

Figure 47. Distribution of individual alpha peak frequencies for the three different baseline conditions (a - c) and the two workload driving conditions (d - e) in Experiment 3.

Figure 47 shows the distribution over subjects of individual alpha frequency peaks in the baseline condition with eyes closed, the two baseline driving conditions that were recorded at the beginning and at the end of a test day, and the two workload driving conditions driven in the morning and in the afternoon. Peak frequencies in each experimental condition were determined by comparing all 11 electrodes and selecting the electrode with maximal power in the alpha frequency range. Again, peak frequencies that were recorded in a relaxed state with eyes closed were clearly distinct from peak frequencies detected during driving (M_{eyes_closed} = 10.38 Hz, SE_{eyes_closed} = 0.16; $M_{baseline\_morning}$ = 9.42 Hz, $SE_{baseline\_morning}$ = 0.22; $M_{baseline\_evening}$ = 9.56 Hz, $SE_{baseline\_evening}$ = 0.20; $M_{listening\_morning}$ = 9.82 Hz, $SE_{listening\_morning}$ = 0.22; $M_{listening\_afternoon}$ = 9.7 Hz, $SE_{listening\_afternoon}$ = 0.23). A repeated-measures ANOVA over 25 drivers' individual peak frequencies with the 5-level main factor "task condition" (baseline with eyes closed vs. driving baseline morning vs. driving baseline afternoon vs. driving and listening task morning vs. driving and listening task afternoon) revealed a significant main effect ($F(2.635, 63.229) = 6.182$, $pGG < .01$, $\eta^2 = .205$). Helmert contrasts showed that the individual alpha frequencies in the resting baseline lay in a statistically significant higher frequency band than the individual alpha frequencies in the four driving conditions ($F(1, 24) = 14.22$, $p < .01$, $\eta^2 = .372$) while none of the subsequent contrasts reached significance indicating that the driving conditions did not differ with respect to peak frequencies (driving baseline morning vs. remaining three conditions: $F(1, 24) = .1.669$, $1 - \beta = .6$ for $\alpha = .05$, $n = 25$, $\eta^2 = .065$, $r = .353$; driving baseline afternoon vs. remaining 2 conditions: $F(1, 24) = 2.036$, $1 - \beta = .939$ for $\alpha = .05$, $n = 25$, $\eta^2 = .078$, $r = .683$; listening task and driving morning vs. listening task and driving afternoon: $F(1, 24) = .1.457$, $1 - \beta = .999$ for $\alpha = .05$, $n = 25$, $\eta^2 = .057$, $r = .937$). In this third experiment, it could be shown that mean resting alpha peak frequencies and mean alpha peak frequencies during driving were clearly distinct and that this even held true when subjects did not perform an additional cognitively demanding task.

Individual alpha peak power amplitudes.

Figure 48 summarizes the mean values for the averaged lateral electrodes in the three baseline and two workload driving conditions. As can be seen, the alpha power during relaxing with eyes closed is strongest over occipital electrodes while the alpha power during the driving baselines is more evenly distributed over all electrode sites with a maximum at parietal electrodes. The alpha power during driving and performing a story listening task is lower than in the three baseline conditions and also evenly distributed over all electrode sites with a maximum at parietal electrodes.

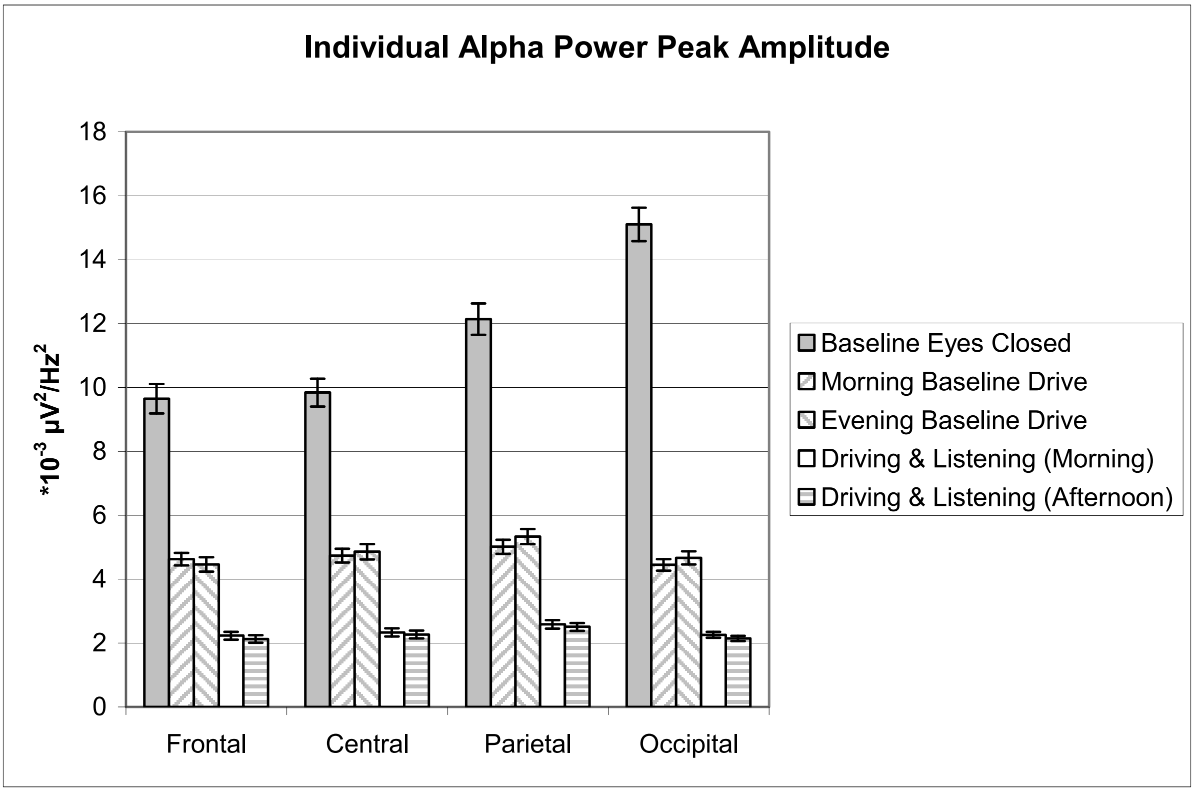

Figure 48. Comparison of individual alpha power recorded during the three experimental baseline conditions (baseline with eyes closed, baseline driving at the beginning and at the end of the experiment) and the two workload driving conditions (morning and afternoon drives). Error bars represent the standard errors.

A repeated-measures ANOVA with the main factors "electrode location" (frontal vs. central vs. parietal and occipital) and "type of task" (baseline with eyes closed vs. morning baseline drive vs. afternoon baseline drive vs. driving and listening task morning vs. driving and listening task afternoon) showed significant effects for the type of task factor($F(1.218, 26.806) = 72.945$, $pGG < .001$, $\eta^2 = .768$) and the electrode location factor ($F(1.658, 36.471) = 18.597$, $pGG < .001$, $\eta^2 = .458$). The interaction between the two main factors was significant ($F(2.029, 44.634) = 29.306$, $pGG < .001$, $\eta^2 = .571$) which is most likely due to the fact that the differences in alpha power between the baseline with eyes closed and the four driving conditions is strongest over occipital electrodes.

Helmert comparisons of the different levels for the type of task factor confirmed the observation that the resting baseline elicited higher power values than any other experimental condition ($F(1, 22) = 72.966$, $p < .001$, $\eta^2 = .768$). Moreover, the contrast between the morning baseline drive and the three other driving conditions was significant ($F(1, 22) = 54.863$, $p < .001$, $\eta^2 = .714$) which is due to the difference between the baseline drives and the two workload driving conditions. Similarly, the afternoon baseline drive was statistically different in alpha power from the two workload driving conditions ($F(1, 22) = 86.692$, $p < .001$, $\eta^2 = .798$), while the two workload drives just failed to reach significance ($F(1, 22) = 3.237$, $p = .086$, $1 - \beta = 1$ for $\alpha =.05$, $n = 23$, $\eta^2 = .128$, $r = .976$). A significant linear contrast for the electrode location factor reflects the strongly increasing alpha power from frontal-central to parietal and finally to occipital electrodes recorded during the baseline with eyes closed ($F(1, 22) = 9.436$, $p < .01$, $\eta^2 = .3$). In sum, the comparison of individual alpha power between resting baseline with eyes closed, the two driving baselines, and the two workload driving conditions showed the clear distinction between the different experimental conditions. In addition to the peak frequencies it could also be demonstrated that the alpha power level is different between baseline driving and driving under high workload, but that each condition was consistent over time of day. Consequently, the power with a range of 2 Hz around the individual and task-specific peak was extracted and statistically analyzed for workload effects during driving. For subjects with no identifiable peak a standard alpha frequency of 10 Hz was used. The same individual alpha frequency window was extracted for driving under high workload due to a secondary task and driving without any additional task.

5.3.2.2 Individual EEG alpha power effect of mental workload during driving.

After the adjustment procedure described in the preceding section, the individual alpha frequency power values were extracted and averaged over subjects. The means over subjects for the workload-relevant comparisons in both recording sessions and the distribution of the effects over the scalp can be seen in *Figure 49*. Performing an additional story listening task led to higher alpha power as opposed to driving only. The effect was broadly distributed over electrodes and stable over recording sessions. Although slightly higher mean values can be observed for the afternoon session, the comparison between sessions in the high workload conditions was not significant (see section 5.3.2.1). However, the replication of the task difficulty effect speaks for its robustness.

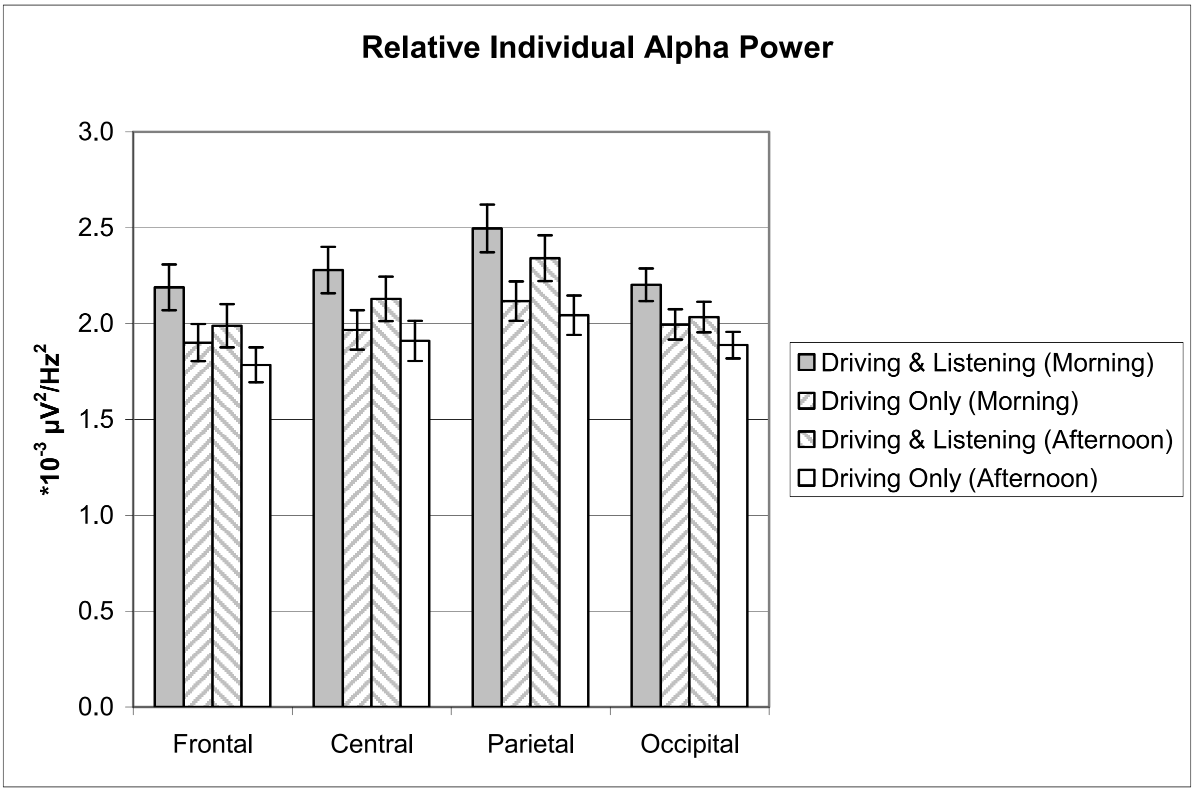

Figure 49. Mean individual EEG alpha power over subjects ($n_{\text{morning session}}$ = 27; $n_{\text{afternoon session}}$ = 26) observed under different levels of mental workload (driving with or without secondary loading task) and for the two different recording sessions (morning vs. afternoon). Error bars represent the standard errors.

For the morning and the afternoon session two separate ANOVAs were run, including the two main factors "electrode location" (frontal vs. central vs. parietal vs. occipital) and "task difficulty" (driving only vs. driving with secondary task). The task difficulty factor revealed a significant effect in the analyses of both recording sessions (morning session: $F(1, 26) = 9.551$, $p < .01$, $\eta^2 = .269$; afternoon session: $F(1, 25) = 6.462$, $p < .05$, $\eta^2 = .205$). For both sessions' analyses the electrode location factor was not significant (morning session: $F(2.396, 62.286) = 2.354$; afternoon session: $F(1.939, 48.47) = 2.883$, $pGG = .067$) which speaks for an evenly distributed effect over all electrodes. For the midline electrode analysis, the task difficulty factor was only significant in the morning session ($F(1, 26) = 5.73$, $p < .05$, $\eta^2 = .181$) and just failed to reach statistical significance in the afternoon ($F(1, 23) = 3.503$, $p = .074$). In sum, in both driving sessions the story listening task revealed significantly higher alpha power and the effect was broadly distributed over all electrode sites.

5.3.3 ECG Indices of Mental Workload

Mean values over all eight blocks were calculated for the heart rate and the three heart rate variability dispersion measures. All mean variables were tested for normal distribution using the Kolmogorov-Smirnov test. Results showed normally distributed mean values for the variables heart rate, SDNN and RMSSD, but not for PNN50. The normally distributed measures were evaluated using repeated-measures ANOVAs including the two main factors “time of day” (morning vs. afternoon session) and “task difficulty” (driving vs. driving with secondary task).

Figure 50 shows the mean heart rate for the eight blocks in the two workload conditions and sessions. A generally higher heart rate can be observed for the afternoon session compared to the morning session as supported by a statistically significant main effect for the time of day factor ($F(1, 25) = 21.747$, $p < .001$, $\eta^2 = .465$). However, mental workload did not have any relevant impact on heart rate ($F < 1$).

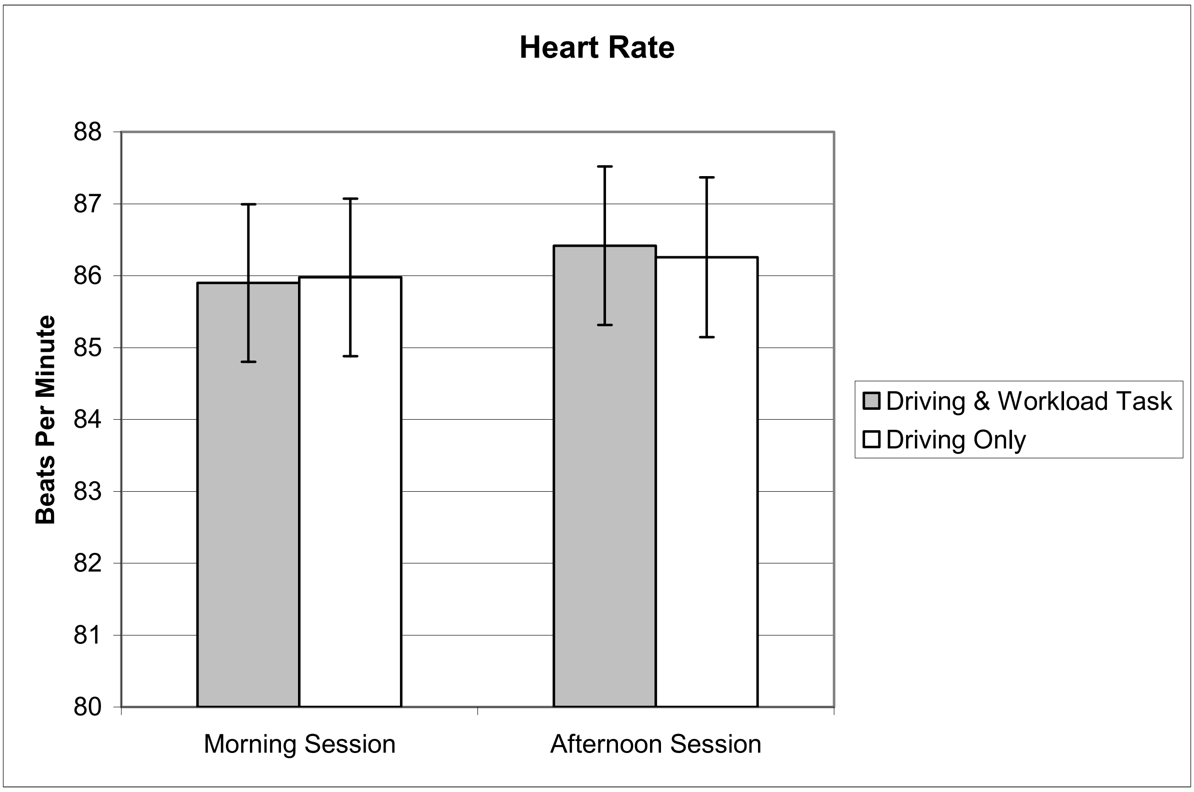

Figure 50. Mean heart rate over subjects (n = 26) for the experimental manipulation of mental workload (driving with secondary task vs. driving without additional task) and for the two recording sessions (morning vs. afternoon). Error bars represent the standard errors.

The statistical analysis for SDNN did not reveal any statistical results for the task difficulty factor ($F(1, 25) = 2.226$) and the time of day factor($F(1,25) = 1.898$). Neither did the analysis for RMSSD show any significance for the task difficulty factor ($F(1, 25) = 3.028$) and time of day factor ($F < 1$).

Figure 51 shows an incoherent pattern of results for the means of the PNN50. As can be seen, the effects of mental workload are minimal. When comparing the driving only conditions, the PNN50 is slightly reduced in the afternoon as opposed to the morning session. The PNN50 data were evaluated using four repeated non-parametric Wilcoxon tests for connected samples to compare the effect of mental workload and time of day. After applying Holm's method to correct for cumulated statistical α due to repeated testing, the only significant result was obtained when comparing the morning and the afternoon session in the "driving only" condition (Z = -.2.781, p < .0125). The means indicated lower PNN50 values for the afternoon session. The comparison of driving only vs. driving and listening task revealed no statistically significant differences in the afternoon (Z = -2.251, p = .024, α(.025) = n.s.) and morning sessions (Z = -.2.139, p = .032, α(.016) = n.s.). In addition, the two high workload conditions did not differ between the morning and the afternoon session (Z = -.825, p = .409, α(.0125) = n.s.).

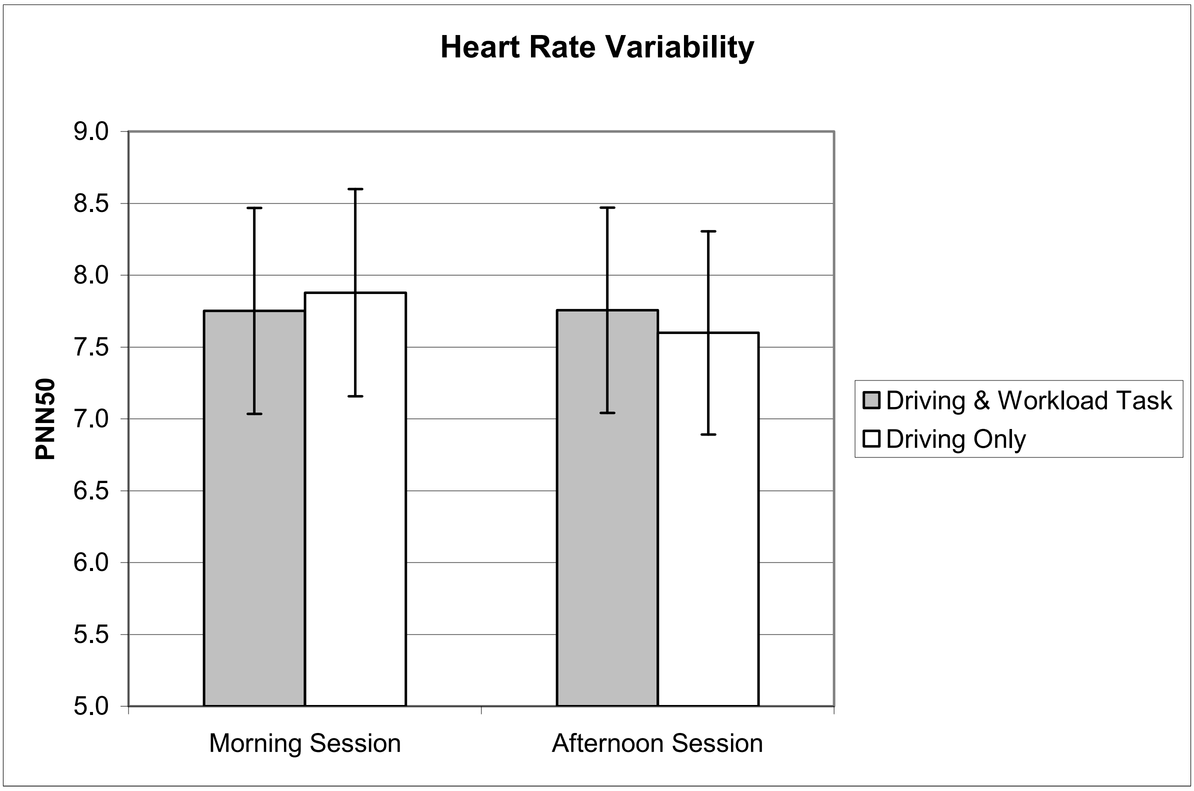

Figure 51. Mean heart rate variability as indicated by the PNN50 over subjects (n = 26) for the experimental manipulation of mental workload (driving with secondary task vs. driving without additional task) and for the two recording sessions (morning vs. afternoon). Error bars represent the standard errors.

In sum, the results of the ECG analysis did not reveal a clear pattern of results with respect to the mental workload manipulation. The only statistically significant effects were obtained for an increase in heart rate and a decrease in heart rate variability (PNN50) for the afternoon session.

5.4 Discussion

The results of Experiment 3 do not only replicate the EEG results of the previous two experiments, but they also represent a replication of the effects within the same experiment over two experimental sessions recorded from the same driver in the morning and afternoon. Moreover, a larger number of subjects (n = 30) was tested and the possibly confounding influence of the surrounding traffic was eliminated by moving the experiment to private terrain and selecting routes that included dirt paths and not public roads. The workload effect of increased EEG alpha power has been robust even despite difficult experimental conditions (e.g. snow). In the following the EEG, ECG, and behavioral data are discussed. This discussion also highlights some general aspects including some limitations of the experimental procedure and the interpretation of the results recorded in this final experiment.

In accordance to Experiments 1 and 2, the combined story listening and word detection task in parallel to driving in Experiment 3 leads to an increase in the drivers' individual EEG alpha power relative to the driving task alone. This effect is consistent for both test session, i.e. for the two routes in the morning and in the afternoon and it is broadly distributed over all electrodes. Drivers really tried hard to follow the content of the stories as confirmed by the analysis of questions about the stories' contents. As in the previous experiments, it has been decided on the basis of an initial screening of the EEG data that an individual and task-specific alpha peak adjustment needed to be performed prior to the analysis. The comparison of the frequency spectra between the baseline recording with eyes closed and the driving condition under the influence of high workload shows characteristic differences in the alpha peak frequencies and amplitude. In addition to the baseline with eyes closed, two baseline driving conditions of 5 - 10 min driving were recorded at the start and end of this experiment. Both driving baselines differ significantly from the baseline with eyes closed. The latter condition shows high alpha power amplitude with a clear occipital distribution and at an upper alpha frequency range. All driving conditions with or without simultaneous cognitive task show lower alpha peak power amplitude and more broadly distributed alpha peak frequencies. The baseline drives reveal higher alpha power than the two drives under high workload, and furthermore both experimental conditions do not differ between morning and afternoon sessions.

Although these results indicate task-specific differences in the alpha frequency band, the increased alpha power during the baseline drives as opposed to the high workload conditions is remarkable. Evidently, the pure driving sections between the 3-min task blocks show a significantly lower alpha power. Based on these observations, it may be concluded that the different driving conditions without additional tasks are not directly comparable to each other. Moreover, results show that alpha power may not only be an indicator of inhibition processes during high mental workload, but that the observed alpha power may also mark another extreme: It can not be excluded that unwanted effects of monotony were recorded in the EEG alpha band. The data pattern resembles some of the respective reports in the literature (see section 2.5.3.1). Reasons for monotony may have lain in major differences between road conditions (broad, tarred road vs. dirt roads) and duration of driving without additional task (5 - 10 min vs. 3-min blocks). Furthermore, drivers were asked to drive silently and the vehicle's governor was set to a maximum speed of 40 km/h. Obviously, once more it could be demonstrated that the functional significance of EEG alpha oscillations is by far more complex than just reflecting one cognitive state.

Other data sources like the ECG failed to show explicit variations due to the mental workload manipulation. Instead the comparison between morning and afternoon recording sessions reveals an increase in heart rate and a decrease in heart rate variability. The latter is indicated by a reduction in the PNN50 when comparing the conditions without additional task presented between the cognitive task blocks. This could demonstrate the drivers' increased effort in the afternoon. The drivers were tested the whole day and due to this extraordinary long time period it is likely that they needed heightened willpower to suppress upcoming fatigue. Nevertheless, due to the missing control measure for breathing, these results again have to be interpreted with some caution (e.g. Respiratory Sinus Arrhythmia). At this point, the comparison for the word detection performance may provide some important supplementation. Lower accuracy and slower reaction times were obtained in the afternoon as opposed to the morning session. These results speak for the interpretation that drivers performed worse in the afternoon, because they were tired.

Besides fatigue there are a few general aspects concerning the experiment that need to be considered. Due to the embedding of drivers into an overall scenario that included both driver and co-driver and that included multiple research questions there might have been an inevitable interference between the tasks of the two participants. For example, the drivers had to slow down at certain parts of the route for the co-drivers to be able to perform their task. In addition, the drivers heard gunfire presented over the headphones at this point. It would be naïve to believe that drivers and co-drivers could perform their tasks completely independent from each other in this particular experimental setting. In fact the overall task situation was quite complex that performance errors, e.g. missing a turn, frequently occurred. Although subjects extensively practiced at the beginning, those errors were unavoidable. Including a tertiary task like in Experiment 2 would have been desirable in order to have an additional performance metric that would be similar over all experimental conditions, but the introduction of such a task would have been unfeasible under these complex circumstances.

In addition, although roads were regularly cleared, adverse weather conditions (ice and snow) resulted in routes being covered with a layer of snow at almost all times. Some drivers were clearly impeded by the snow and this made it necessary to instruct them to drive at a slower speed than initially intended when planning the routes. Even after shortening the routes, drivers were occupied longer than planned in the beginning (30 - 45 min as opposed to 20 - 30 min). This produced 15 - 20 min long sections at the end of each round without stimulus presentation for the driver that could not be shortened due to the co-drivers' tasks that depended on passing certain sections of the route.

For these various reasons it has to be concluded that an investigation only focusing on the drivers might have allowed a more economical experimental design possibly resulting in a more coherent picture of results. For example, such a design could have included only a single route and a tertiary task running over all experimental conditions. Nevertheless and despite the objections discussed above, definite and positive results for the EEG alpha band were obtained in this final driving experiment confirming the hypothesis of endogenous distraction based on mental workload induced by a secondary task. The results are consistent with the preceding two experiments and they round off the overall picture of experimental results by their replication.

6 Overall discussion and outlook

The present work examined changes in the EEG alpha band of mentally strained car drivers. Mental workload was induced by auditory secondary tasks during which drivers had to listen carefully to the contents of presented audio books. These narrations created a type of workload that was similar to an intensive conversation with a passenger in the car or similar to a talk on the phone. A total of three experiments were run to investigate reliability and generalizability of the effects in the EEG as well as to examine different aspects of the drivers' workload. In doing so a power increase in the individual EEG alpha band could be demonstrated with increased mental workload which was stable over single as well as dual-task conditions. This effect supports the assumption of a top-down controlled concentration onto the auditory task. It is concluded that the alpha band can provide an index of the drivers' suppression of conscious processing of visual stimuli stemming from the primary driving task to attain an optimized processing of the secondary task.

The laboratory results build the foundation for Experiment 2, which replicate this effect of individual alpha power increase under real-traffic driving conditions on a German freeway. The findings reveal that, a second newly introduced mental arithmetic task did not show the same alpha band effects as the listening task. However, by using a simple stimulus-response tertiary task the workload manipulation due to the mental arithmetic task was shown in the ERPs elicited by the stimuli. It was also demonstrated that the mental arithmetic task led to an increase in the N1 amplitude which is in accordance with the widespread view that this component is related to increased attention. Moreover, a reduction in the P3 was shown in this secondary task condition which replicated already known effects (for review see Manzey, 1998). Besides the increase in the individual alpha band, a latency delay of about 50 ms in the N1 as well as a related significant reduction in the

N1 amplitude was detected for the listening task. It is assumed that this latency delay reflects a delayed processing of concurrent auditory stimuli.

In a third experiment, the effect of alpha power increase under high mental workload was once more replicated in an additional field experiment. This effect was even demonstrated to be stable over two different measurement times (morning, afternoon). The drivers' task was identical to Experiment 1 and very similar to the one used in Experiment 2, but this time a larger subject population was tested and the experiment was run on dirt roads under the exclusion of any traffic. In all three experiments and besides the EEG data, additional behavioral and electro-cardiac measures were collected which provided an additional validation for the EEG data in terms of a multi-level account (Manzey, 1998). The clear results in all three experiments speak for the high reliability of the measurement and the generalizability over different experimental settings of the observed effects.

With respect to the already existing literature on the topic of mental workload during driving it can be concluded that the present EEG investigation examined a completely new aspect. In general, EEG studies on driving are rare. Most of the studies were run in the laboratory and in connection with driving simulations. Up to now, only one study is known that investigated mental workload in real traffic driving by using EEG (Schmidt, 2006). However, in this work mental workload was induced by driving maneuvers which involved a relatively high contamination of the EEG data with muscle artifacts. In this case such a contamination could hardly be avoided since the workload manipulation was indistinguishably connected to increased muscle activity. In contrast, the present work holds the advantage that mental workload was induced by secondary task presented in a strict block design. This elicits a special kind of driver workload which also differs qualitatively from the workload due to demanding driving situations. While demanding driving situations mostly require drivers' visual attention that is predominantly directed onto the road and the surrounding traffic, attention to the secondary tasks that were used here is mostly internalized and without a visual component. In this context, Recarte and Nunes (2003) coined the term of “endogenous distraction” of the driver. In their real-traffic driving experiments they systematically induced mental workload by involving the drivers in conversations with a passenger or a person on the phone. By analyzing eye movements they demonstrated that strained drivers had a spatial gaze distribution that involved worse perception of stimuli in their visual periphery. Strayer and Drews (2007) tested subjects in a driving simulator and used the P300 ERP component to show that drivers that were under the influence of a conversation made more mistakes in encoding warning lights. Based on the observation of “inattentional blindness”, the authors called attention to a potentially increased accident risk in conversational situations. The results of the present work closely relate to these studies by showing that the impact of this kind of mental workload is reflected by an increase of the power in the EEG alpha band. At the same time a bridge is built between investigations in traffic psychology on the one side and results from pure laboratory research on the topic of “internalized attention” on the other side. These previous laboratory studies demonstrated that internalized as opposed to externalized attention as well as increased task difficulty led to an increase in alpha power (e.g. Cooper, Croft et al., 2003, Jensen et al., 2002; Ray & Cole, 1985b). Generally, this increase has been attributed to an inhibition of task-irrelevant cortical areas. This situation can be very well transferred to the car driving situation. For example, drivers who are currently listening to an important emergency message on the radio have to shift their focus of attention away from the driving task and onto the listening task. In this way, it seems by all means plausible that drivers try to suppress visual distractors that could distract them from listening. Proof for this assumption may easily be observed from the drivers’ behavior itself. For example, many drivers usually change to the slower right highway lane when receiving a call on their car phone. It is evident, that with this behavior drivers try to reduce the predominantly visual demands from the driving situation to be able to allocate more attentional resources onto the conversation. The cortical effects of intensive internalized attention can be observed in the alpha band. The results of the present three experiments confirm the correctness of transferring insights from the laboratory to the everyday situation inside the vehicle.

Nevertheless, at this point, there is also some caution advised if one wants to suggest a simplistic function of the alpha band as a workload indicator based on these results only. There is also a great amount of studies that showed the exact opposite impact of mental workload. On closer inspection, it seems that particularly the type of task (internalized vs. externalized attention and a great proportion of mental visualization in working memory involved) and the task situation (single vs. multiple tasks) play a decisive role for the direction of the effects in the alpha band. In this way, Jensen and colleagues (2002) argued that alpha desynchronization due to an n-back task as demonstrated by Gevins and colleagues (Gevins, Smith, McEvoy, & Yu, 1997) can be best explained by a visual processing strategy when subjects performed the task: „Assuming that alpha activity reflects suppression of processing in visual areas, increased demands to the visual system with memory load would explain the posterior decrease in alpha activity“ (p. 881). Moreover, there is demonstrably a strong impact of a person's general activation state onto the alpha band so that above all an increase in the alpha band may also be rated as a sign of monotony or fatigue. These multiple interconnections which have not yet been sufficiently clarified make it a difficult task to examine the alpha band even if extensive planning of an investigation was done. Due to increased expenditure of time given by the time-consuming preparation of the EEG electrodes, confounding variables such as fatigue are hardly avoidable. In Experiment 3 the impact of these types of confounding variables could even be demonstrated in detail with the help of extra driving baselines. Despite explicit characteristic differences in the alpha band between resting baseline with eyes closed and alpha band effects during driving, the 5 - 10 min baseline drives show a clear increase in the alpha band which is above the alpha power when driving under high workload and which can only be explained by heightened monotony during the baseline drives. Nevertheless, the same subjects show clear workload effects when comparing task blocks as opposed to driving sections in-between these blocks without task. Additional research is definitely necessary to further resolve the complexity of the functional significance of the alpha band. In other words, the existing alpha band theories need to be summarized and differentiated in more detail. General activation theories are too unspecific and inhibition theories need more empirical support.

An arrear of research is also apparent in the consideration of inter-individual differences in the data analysis. In this case no standard procedure has been established so far. For the present experiments an individual adjustment of the alpha frequency band was indispensable to attain a satisfactory signal-to-noise ratio. But for specific reasons there are also contrary opinions arguing against such a procedure. For example, Berka and colleagues (2007) recently stated: „Although individualized models can provide highly accurate classifications of cognitive state, they may be impractical in operational environments due to the additional time required for personnel to obtain the model data and the computer processing time and capacity required to create complex models such as those required by ANNs [artificial neural networks]“ (p. 232). This argument may be true for specific application contexts (e.g. military operations). Many authors arguing in this way hence employ a bottom-up approach in which they integrate multiple data sources and assign the collected data to different cognitive states using sophisticated classification procedures. The successes that they achieved so far have been considerable. However after all, such an application-oriented practice can only sparsely contribute to answering the question about the originally occurring cognitive mechanisms and their neural correlates in the brain. For such a purpose, the closer examination of a single mechanism such as the inhibition mechanism in the alpha band seems more promising. Future works can build on these results and for example establish an individual calibration task for the alpha band as a standard measurement. This should allow a better differentiation and prediction of the effects and make it conceivable to use shorter time windows to determine cognitive states in the near future.

Another important point , that may be of interest especially for basic research concerns the assumption of a ring of inhibition postulated by several authors (e.g. Klimesch, 1997; Pfurtscheller, 2003). They assume that a cortical region with active neural processing is marked by a suppression of alpha power. At the same time surrounding, task-irrelevant areas are actively inhibited which is reflected by an increase of the power in the alpha band in these regions. The present three experiments examined only a small number of 11 electrodes and the validation of such a ring of inhibition as for example done by Cooper, Croft et al. (2003) was therefore not possible. Future studies could include more electrodes to find additional support for this assumption.

For research and development in traffic psychology after all, there is one main goal lying behind measuring mental workload with EEG which is to detect states of overload of the driver and to initiate effective countermeasures. To guarantee safety in road traffic this has to be reliably performed within seconds. Among known brain imaging methods the EEG is the one tool with the highest time resolution making it the most suitable candidate for this task. The EEG can be specifically used to evaluate and advance the effectiveness auf driver assistance systems. For example, a straightforward approach is to use changes in the ERPs in an artificially introduced oddball-task to evaluate changes in drivers' cognitive states (e.g. Hagemann, Schrauf, & Kincses, 2004).

However, such an account always requires averaging over a longer time period. For example, using an oddball paradigm the impact of a driver assistance system such as an adaptive cruise control (ACC) can be easily tested since this system is mostly active over long time periods. On the other hand, the rare interception of some other assistance systems makes it difficult to employ a systematic experimental assessment of their impact onto the driver by using the oddball task. Moreover, the oddball approach requires an extra task. At this point, the alpha band provides the possibility to measure independently from any stimulus. While the potential for advances with the oddball paradigm seems to be depleted after decades of research, an alpha band analysis with high temporal resolution appears to be near at hand with the help of some more research. Thereby, it becomes conceivable to analyze the cognitive impact of different human-machine-interfaces onto the driver. In order to achieve reasonable applications within this area, a strong interconnection between suitable experimental designs and advanced analysis methods has to be attained. Moreover, appropriate neuroscientific knowledge is needed to formulate the right hypotheses and to allow a meaningful interpretation of the results. For this purpose, the findings of the present work make an important contribution by transferring a practically relevant, electrophysiological phenomenon from the laboratory to real road traffic. The presented results leave room for optimism for future EEG investigations of ergonomic research questions in the field of driver-vehicle-interaction.

7 References

Alm, H., & Nilsson, L. (1994). Changes in driver behaviour as a function of hands free mobile phones – a simulator study. *Accident analysis and prevention*, 26, 441-451.

Alm, H., & Nilsson, L. (1995). The effects of a mobile telephone task on driver behaviour in a car following situation. *Accident analysis and prevention*, 27, 707-715.

Althaus, M., Mulder, L.J.M., Mulder, G., Van Roon, A.M., & Minderaa, R.B. (1998), Influence of respiratory activity on the cardiac response pattern to mental effort. *Psychophysiology*, 35, 420-430.

Angell, L., Auflick, J., Austria, P.A., Kochhar, D., Tijerina, L., BIever, W., et al. (2006). *Driver workload metrics task 2 final report.* Springfield. (NTIS No. DOT HS 810635).

Backs, R.W., & Seljos, K.A. (1994).Metabolic and cardiorespiratory measures of mental effort: The effects of level of difficulty in a working memory task. *International Journal of Psychophysiology*, 16, 57-68.

Backs, R.W., Ryan, A.M., & Wilson, G. F. (1994). Psychophysiological measures of workload during continuous manual performance. *Human Factors*, 36, 514-531.

Bamberg, U. (2002). Normung zur psychischen Belastung - aus Sicht der Arbeitnehmer. *DIN-Mitteilung*, 81, 529-533.

Bartenwerfer, H. (1969). Einige praktische Konsequenzen aus der Aktivierungstheorie. *Zeitschrift für experimentelle und angewandte Psychologie,* 16, 195-222.

Baumgartner, N. (2004). *Der Effekt von Polarität der Textdarstellung und Umgebungsbeleuchtung auf die Performanz und psychophysiologische Beanspruchung bei Bildschirmarbeit*. Doctoral dissertation, Heinrich-Heine University Düsseldorf, Germany, 2004 (URN No. urn:nbn:de:hbz:061-20040713-000874-7). Retrieved February 21, 2008, from http://docserv.uni-duesseldorf.de/servlets/DocumentServlet?id=2874

Beede, K.E., & Kass, S.J. (2006). Engrossed in conversation: the impact of cell phones on simulated driving performance. *Accident analysis and prevention*, 38, 415-421.

Benda, H. von, Hoyos, C. Graf, & Schaible-Rapp, A. (1983). *Klassifikation und Gefährlichkeit von Straßensituationen. Bericht zum Forschungsprojekt 7320*. Bergisch Gladbach: Bundesanstalt für Straßenwesen.

Berg, P. & Scherg, M. (1991). Dipole models of eye movements and blinks. *Electroencephalography and Clinical Neurophysiology*, 79, 36-44.

Berger, H. (1929). Über das Elektrenzephalogramm des Menschen. *Archives of Psychiatry*, 87, 527-570.

Berka, C., Levendowski, D.J., Lumicao, M.N., Yau, A., Davis,G., Zivkovic, V.T., et al. (2007). EEG Correlates of Task Engagement and Mental Workload in Vigilance, Learning, and Memory Tasks. *Aviation, Space, and Environmental Medicine*, 78, 231-244.

Bernotat, R., & Käppler, W.B. (1985). Wirksystem Fahrer-Fahrzeug-Umwelt. In K. Rompe (Ed.), *Verkehrssicherheit und Wirksystem Fahrer-Fahrzeug-Umwelt* (pp. 13-44). Köln: Verlag TÜV-Rheinland.

Birbaumer, N., & Schmidt, R.F. (1996). *Biologische Psychologie* (3rd ed.). Heidelberg: Springer-Verlag.

Birbaumer, N., & Schmidt, R.F. (2006). *Biologische Psychologie* (6th ed.). Heidelberg: Springer Medizin Verlag.

Bösel, R. (1996). *Die EEG-Grundaktivität. Ein Laborhelfer.* Regensburg: Roderer Verlag.Bornemann, E. (1942). Untersuchungen über den Grad der geistigen Beanspruchung. *Arbeitsphysiologie*, 12, 142-191.

Bouyer, J.J., Montaron, M.F., Fabre-Thorpe, N., & Rougeul, A. (1986). Compulsive attentive behaviour after lesions of the ventral striatum in the cat. *Experimental Neurology*, 92, 698-712.

Brookhuis, K. A., & De Waard, D. (1993). The use of psychophysiology to assess driver status. *Ergonomics*, 36, 1099 – 1110.

Brookhuis, K.A., & De Waard, D. (2001). Assessment of drivers' workload performance and subjective and physiological indexes. In P.A. Hancock & P.A. Desmond (Eds.), *Stress, Workload and Fatigue* (pp. 321-333). London: Lawrence Erlbaum.

Brookhuis, K.A., De Vries, G., & De Waard, D. (1991). The effects of mobile telephoning on driving performance. *Accident analysis and prevention*, 23, 309-316.

Brookhuis, K.A., Louwerens, J.W., & O'Hanlon, J.F. (1986). EEG energy-spectra and driving performance under the influence of some antidepressant drugs. In J.F. O'Hanlon & J.J. de Gier (Eds.), *Drugs and driving* (pp. 213-221). London: Taylor and Francis.

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

Brown, I.D., Tickner, A.H., & Simmonds, D.C.V. (1969). Interference between concurrent tasks of driving and telephoning. *Journal of Applied Psychology*, 53, 419-424.

Brunia, C.H.M. (1989). Correcting ocular artifacts in the EEG: a comparison of several methods. *Journal of Psychophysiology*, 3, 1-50.

Busch, N.A., & Herrmann, C.S. (2003). Object-load and feature-load modulate EEG in a short-term memory task. *Neuroreport*, 14, 1721-1724.

Caird, J.K., Scialfa, C.T., Ho, G., & Smiley, A. (2004). Effects of cellular telephones on driving behaviour and crash risk: results of meta-analysis. Retrieved August 5, 2007, from http://www.ama.ab.ca/images/images_pdf/FinalReport_CellPhones4.pdf

Calhoun, V.D., Pekar, J.J. McGinty, V.B., Adali, T., Watson, T.D., & Godfrey, G.D. (2002). Different activation dynamics in multiple neural systems during simulated driving. *Human Brain Mapping*, m16, 158-167.

Carroll, D., Turner, J.R., & Hellawell, J.C. (1986). Heart rate and oxygen consumption during active psychological challenge: The effects of level of difficulty. *Psychophysiology*, 23, 174-181.

Casali, J.G., & Wierwille, W.W. (1983). A comparison of rating scale, secondary task, physiological, and primary-task workload estimation techniques in a simulated flight task emphasizing communications load. *Human Factors*, 25, 623-641.

Caton, R. (1875). The electric currents of the brain. *British Medical Journal*, 2, 278.

Cerezuela, G.P., Tejero, P., Chóliz, M., Chisvert, M., & Monteagudo, M.J. (2004). Wertheim's hypothesis on 'highway hypnosis': empirical evidence from a study on motorway and conventional road driving. *Accident Analysis and Prevention*, 36, 1045-1054.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Earlbaum Associates.

Coles, M.G.H., & Rugg, M.D. (1997). Event-related brain potentials: an introduction. In M.G.H. Coles & M.D. Rugg (Eds.), *Electrophysiology of mind. Event-related brain potentials and cognition* (pp. 1-26). Oxford: Oxford University Press.

Coles, M.G.H., Smid, H.G.O.M., Scheffers, M.K., & Otten, L.J. (1997). Mental chronometry and the study of human information processing. In M.G.H. Coles & M.D. Rugg (Eds.), *Electrophysiology of mind. Event-related brain potentials and cognition* (pp. 86-131). Oxford: Oxford University Press.

Consiglio, W., Driscoll, P., Witte, M., & Berg, W.P. (2003). Effect of cellular telephone conversations and other potential interference on reaction time in a braking response. *Accident analysis and prevention*, 35, 495-500.

Controller Area Network. (2007). In *Wikipedia, The Free Encyclopedia*. Retrieved May 25, 2007, from http://en.wikipedia.org/w/index.php?title=Controller_Area_Network&oldid=132459092

Cooper, N.R., Burgess, A.P., Croft, R.J., & Gruzelier, J.H. (2006). Investigating evoked and induced electroencephalogram activity in task-related alpha power increases during an internally directed attention task. *Neuroreport*, 17, 205-208.

Cooper, N.R., Croft, R.J., Dominey, S.J.J., Burgess, A.P., & Gruzelier, J.H. (2003). Paradox lost? Exploring the role of alpha oscillations during externally vs. internally directed attention and the implications for idling and inhibition hypotheses. *International Journal of Psychophysiology*, 47, 65-74.

Cooper, P.J., Zheng, Y., Richard, C., Vavrik, J., Heinrichs, B., & Siegmund, G. (2003). The impact of hands-free message reception/response on driving task performance. *Accident analysis and prevention*, 35, 23-35.

Courchesne, E., Hillyard, S.A., & Galambos, R. (1975). Stimulus novelty, task relevance and the visual evoked potential in man. *Electroencephalography and Clinical Neurophysiology*, 39, 131-143.

De Waard, D. & Brookhuis, K.A. (1991). Assessing driver status: a demonstration experiment on the road. *Accident and analysis prevention*, 23, 297-307.

De Waard, D. (1996). *The measurement of drivers' mental workload*. Doctoral dissertation, University of Groningen, Haren, The Netherlands. Retrieved November 1, 2007, from http://irs.ub.rug.nl/ppn/297329391

De Waard, D. (2002). Mental Workload. In R. Fuller & J.A. Santos (Eds.), *Human Factors for Highway Engineers* (pp. 161-175). London: Elsevier.

Delorme, A., & Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics, *Journal of Neuroscience Methods,* 134, 9-21.

Delorme, A., Makeig, S., Humphries, C., Enghoff, S., Jung, T.-P., Bell, T. et al. (2002). *EEGLAB (Version 5.03)*. University of California San Diego.

Deutsches Institut für Normung e.V. (1987). *DIN 33 405. Psychische Belastung und Beanspruchung; Allgemeines, Begriffe.* Berlin: Beuth Verlag.

Deutsches Institut für Normung e.V. (2000). *DIN EN ISO 10075-1:2000. Ergonomische Grundlagen bezüglich psychischer Arbeitsbelastung. Teil 1: Allgemeines und Begriffe.* Berlin: Beuth Verlag.

Dien, J., Spencer, K.M., & Donchin, E. (2003). Localization of the event-related potential novelty response as defined by principial component analysis. *Cognitive Brain Research*, 17, 637-650.

Donchin, E., & Coles, M.G.H. (1988). Is the P300 component a manifestation of context updating? *Behavioral and Brain Sciences*, 11, 357-374.

Donges, E. (1982). *Aspekte der aktiven Sicherheit bei der Führung von Personenkraftwagen.* Automobil-Industrie, 2, 183-190.

Donges, E., & Naab, K. (1996). Regelsysteme zur Fahrzeugführung und -stabilisierung in der Automobiltechnik. *Automatisierungstechnik*, 44, 226-236.

Duncan, J., Williams, P., & Brown, I. (1991).Components of driving skill: experience does not mean expertise. *Ergonomics*, 34, 919-937.

Duncan-Johnson, C.C., & Donchin, E. (1977). On quantifying surprise: The variation in event-related potentials with subjective probability. *Psychophysiology*, 14, 456-467.

Duncan-Johnson, C.C., & Donchin, E. (1983). The P300 component of the event-related brain potential as an index of information processing. *Biological Psychology*, 14, 1-52.

Eggemeier, F.T., & Wilson, G.F. (1991). Performance-based and subjective assessment of workload in multi-task environments. In D.L. Damos (Ed.), *Multiple task performance* (pp. 217-278). London: Taylor & Francis.

Engenuity Technologies Inc. (2004). *Virtual Application Prototyping System (VAPS)* [computer software]. Montreal, CA.

Fairclough, S.H. (1991). *Adapting the TLX to measure driver mental workload.* DRIVE Project V1017 (BERTIE) Report No. 71. Loughborough: HUSAT Research Institute.

Fairclough, S.H., Venables, L., & Tattersall, A. (2005). The influence of task demand and learning on the psychophysiological response. *International Journal of Psychophysiology*, 56, 171-184.

Färber, B., & Färber, B. (2004). Mehr Verkehrssicherheit durch intelligente Steuerung von Telematik-Systemen? In B. Schlag (Ed.), *Verkehrspsychologie* (pp. 317-333). Lengerich, Pabst.

Fastenmeier, W. (1995). *Autofahrer und Verkehrssituation. Neue Wege zur Bewertung von Sicherheit und Zuverlässigkeit moderner Sicherheitssysteme.* Köln: Verlag TÜV Rheinland.

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, 39, 175-191.

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, 129-145.

Frieling, E., & Hoyos, C. Graf (1978). *Fragebogen zur Arbeitsanalyse (FAA)*. Bern: Huber.

Gaillard, A.W.K. (1993). Comparing the concepts of mental load and stress. *Ergonomics*, 36, 991-1005.

Gelau, C. (2004). Fahrerablenkung durch Informations- und Kommunikationssysteme im Fahrzeug: Auswirkungen auf das Fahrerverhalten und die Verkehrssicherheit. In B. Schlag (Ed.), *Verkehrspsychologie* (pp. 297-316). Lengerich, Pabst.

Gevins, A. & 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., 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*, 7, 374-385.

Gibson, J.J., & Crooks, L.E. (1938). A theoretical field analysis of automobile driving. *The American Journal of Psychology*, 51, 453-471.

Gimeno, P.T., Cerezuela, G.P., & Montaés, M.C. (2006). On the concept and measurement of driver drowsiness, fatigue and inattention: implications for countermeasures. *International Journal of Vehicle Design*, 42, 67-86.

Glass, A. (1966). Comparison of the effect of hard and easy mental arithmetic upon blocking of the occipital alpha rhythm. *Quarterly Journal of Experimental Psychology*, 8, 142-152.

Goodman, M.J., Tijerina, L., Bents, F.D., & Wierwille, W.W. (1999). Using cellular telephones in vehicles: safe or unsafe? *Transportation Human Factors*, 1, 3-42.

Gratton, G., Coles, M.G.H., & Donchin, E. (1983). A new method for off-line removal of ocular artifacts. *Electroencephalography and Clinical Neurophysiology*, 55, 468-481.

Green, P. (2004). Driver distraction, telematics design, and workload managers: saftey issues and solutions. *Proceedings of the 2004 international congress on transportation electronics* (*convergence 2004, SAE publication P-387)*, 165-180. Warendale, PA: Society of automotive engineers. Retrieved February, 18, 2008, from http://www.umich.edu/~driving/publications/GreenConvergence04paper4b.pdf

Groeger, J.A. (2000). *Understanding driving: applying cognitive psychology to a complex everyday task.* Philadelphia: Taylor & Francis.

Groeger, J.A. (2002). Trafficking in cognition: applying cognitive psychology to driving. *Transportation Research Part F*, 5, 235-248.

Groeger, J.A., & Clegg, B.A. (1997). Automaticity and driving. Time to change a gear conceptually. In J.A. Rothengatter & E.C. Vaya (Eds.), *Traffic and transport psychology: Theory and Application* (pp. 137-146). Amsterdam: Elsevier.

Gruzelier, J., Liddiard, D., Davies, L., & Wilson, L. (1990). Topographical EEG differences between schizophrenic patients and controls during neuropsychological functional activation. *International Journal of Psychophysiology*, 8, 275-282.

Gstalter, H., & Fastenmeier, W. (1995). Auswirkungenvon Navigationsinformationen im Kraftfahrzeug: Mögliche Sicherheitseffekte, wichtige Fahrervariablen und Folgerungen für empirische Untersuchungen. In Fastenmeier (Ed.), *Autofahrer und Verkehrssituation. Neue Wege zur Bewertung von Sicherheit und Zuverlässigkeit moderner Sicherheitssysteme* (pp. 79-95). Köln: Verlag TÜV Rheinland.

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

Hagemann, K. (2002). *Electrophysiological evidence for hemispheric asymmetry in category ambiguity resolution.* Unpublished Diplomarbeit, University of Potsdam, Germany.

Hagemann, K., Schrauf, M., & Kincses, W. (2004). Measuring driver's cognitive workload using ERP. *Journal of Psychophysiology*, 18, 232.

Hale, A.R., Stoop, J., & Hommels, J. (1990). Human error models as predictors of accident scenarios for designers in road transport systems. *Ergonomics*, 33, 1377-1388.

Halgren, E., Boujon, C., Clarke, J., Wang, C., & Chauvel, P. (2002). Rapid distributed fronto-parieto-occipital processing stages during working memory in humans. *Cerebral Cortex*, 12, 710-728.

Harbluk, J.L., Noy,Y.I., Trbovich, P.L., & Eizenman, M. (2007). An on-road assessment of cognitive distraction: Impacts on drivers' visual behavior and braking performance. *Accident Analysis and Prevention*, 39, 372-379.

Hari, R., Salmelin, R., Mäkelä, J.P., Salenius, S., & Helle, M. (1997). Magnetoencephalographic cortical rhythms. *International Journal of Psychophysiology*, 26, 51-62.

Harms, L. (1991). Variation in drivers' cognitive load. Effects of driving through village areas and rural junctions. *Ergonomics*, 34, 151-160.

Harrer, H. (1952). *Sieben Jahre in Tibet.* [Recorded by M. M. Schwarz, CD]. Marburg: Verlag und Studio für Hörbuchproduktionen.

Hart, S.G., & Staveland, L.E. (1988). Development of the NASA-TLS (Task Load Index): Results of empirical and theoretical research. In P.A. Hancock & N. Meshkati (Eds.), *Human mental workload*. Amsterdam: North Holland.

Hebb, D.O. (1949). *The organization of behavior.* New York: Wiley.

Hebert, R., & Lehmann, D. (1977). Theta bursts: An EEG pattern in normal subjects practising the transcendental meditation technique. *Electroencephalography and Clinical Neurophysiology*, 42, 397-405.

Herbert, A.M., Gordon, G.E., & McCulloch, D.L. (1998). A "passive" event-related potential?". *International Journal of Psychophysiology*, 28, 11-21.

Hering, K. (1999). *Situationsabhängiges Verfahren zur standardisierten Messung der kognitiven Beanspruchung im Straßenverkehr. Literaturübersicht und empirische Felduntersuchung.* Doctoral dissertation, University Köln, Germany, 1999 (URN No. urn:nbn:de:hbz:38-5145). Retrieved February 21, 2008, from http://deposit.d-nb.de/cgi-bin/dokserv?idn=958775532

Herrmann, C.S., & Demiralp, T. (2005). Human EEG gamma oscillations in neuropsychiatric disorders. *Clinical Neurophysiology*, 116, 2719-2733.

Hill, S.G., Iavecchia, H.P., Byers, J.C., Bittner, A.C., Jr., Zaklad, A.L., & Christ, R.E. (1992). Comparsion of four subjective workload rating scales. *Human Factors*, 34, 429-440.

Hillyard, S.A., & Galambos, R. (1970). Eye movement artifact in the CNV. *Electroencephalography and Clinical Neurophysiology*, 28, 173-182.

Hillyard, S.A., Hink, R.F., Schwent, V.L., & Picton, T.W. (1973). Electrical signs of selective attention in the human brain. *Science*, 182, 177-179.

Hillyard, S.A., Vogel, E.K., & Luck, S.J. (1998). Sensory gain control (amplification) as a mechanism of selective attention: Electrophysiological and neuroimaging evidence. *Philosophical Transactions of the Royal Society: Biological Sciences*, 353, 1257-1270.

Hinton, S. C., & Meck, W. H. (1997). The 'internal clocks' of circadian and interval timing. *Endeavour*, 21, 82–87.

Hockey, G.R.J. (1997). Compensatory control in the regulation of human performance under stress and high workload: A cognitive-energetical framework. *Biological Psychology*, 45, 73-93.

Hoffmann, J.E. (1990). Event-related potentials and automatic and controlled processes. In J.W. Rohrbaugh, R. Parasuraman & R. Jr. Johnson (Eds.), *Event related brain potentials* (pp. 145-157). New York: Oxford University Press.

Holcombe, V., Sterman, M.B., Goodman, S.J., & Fairchild, M.D. (1979). The immobilization response in rhesus monkey: A behavioral and electroencephalography study. *Experimental Neurology*, 63, 420-435.

Hole, G. (2007). *The Psychology of Driving.* Mahwah: Lawrence Erlbaum Associates.

Hopf, J.-M., Vogel, E.K., Woodman, G.F., Heinze, H.-J., & Luck, S.J. (2002). Localizing visual discrimination processes in time and space. *Journal of Neurophysiology*, 88, 2088-2095.

Horikawa, E., Okamura, N., Tashiro, M., Sakurada, Y., Maruyama, M., Arai, H. et al. (2005). The neural correlates of driving performance identified using positron emission tomography. *Brain and cognition*, 58, 166-171.

Hoyos, C. Graf, Fastenmeier, W., & Gstalter, H. (1995). Forderungen an eine verhaltensorientierte Verkehrssicherheitsarbeit. In Fastenmeier (Ed.), *Autofahrer und Verkehrssituation. Neue Wege zur Bewertung von Sicherheit und Zuverlässigkeit moderner Sicherheitssysteme* (pp. 11-26). Köln: Verlag TÜV Rheinland.

International Organization for Standardization (1991). *ISO 10075:1991(E). Ergonomic principles related to mental work-load - General terms and definitions.* Geneva, Switzerland: International Organization for Standardization.

International Organization for Standardization (1996). *ISO 10075-2:1996(E). Ergonomic principles related to mental work-load - Part 2: Design principles.* Geneva, Switzerland: International Organization for Standardization.

International Organization for Standardization (2004). *ISO 10075-3:2004(E). Ergonomic principles related to mental work-load - Principles and requirements concerning methods for measuring and assessing mental workload.* Geneva, Switzerland: International Organization for Standardization.

Isreal, J., Chesney, G., Wickens, C., & Donchin, E. (1980). P300 and tracking difficulty: Evidence for multiple resources in dual-task performance. *Psychophysiology*, 17, 259-273.

Jahn, G., Oehme, A., Krems, J.F., & Gelau, C. (2005). Peripheral detection as a workload measure in driving: effect of traffic complexity and route guidance system use in a driving study. *Transportation research part F*, 8, 255-275.

Jasper, H.H. (1958). The ten-twenty electrode system of the International Federation. *Electroencephalography and Clinical Neurophysiology,* 10, 371-375.

Jensen, O., Gelfand, J., Kounios, J., & Lisman, J.E. (2002). Oscillations in the alpha band (9-12 Hz) increase with memory load during retention in a short-term memory task. *Cerebral Cortex,* 12, 877-882.

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

Johannsen, G. (1993). *Mensch-Maschine-Systeme.* Berlin: Springer Verlag.

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

Joyce, C.A., Gorodnitsky, I.F., & Kutas, M. (2004). Automatic removal of eye movement and blink artifacts from EEG data using blind component separation. *Psychophysiology*, 41, 313-325.

Jung, T.P., Makeig, S., Humphries, C., Lee, T.W., McKeown, M.J., Iragui, V., et al. (2000). Removing electroencephalographic artifacts by blind source separation. *Psychophysiology*, 37, 163-178.

Jung, T.P., Makeig, S., Westerfield, M., Townsend, J., Courchesne, E., & Sejnowski, T.J. (2000). Removal of eye activity artifacts from visual event-related potentials in normal and clinical subjects. *Clinical Neurophysiology*, 111, 1745-1758.

Just, M.A., Carpenter, P.A., Keller, T.A., Emery, L., Zajac, H., & Thulborn, K.R. (2001). Interdependence of nonoverlapping cortical systems in dual cognitive tasks. *Neuroimage*, 14, 417-426.

Kahana, M.J., Sekuler, R., Caplan, J.B., Kirschen, M. & Madsen, J.R. (1999).Human theta oscillations exhibit task dependence during virtual maze navigation. *Nature*, 399, 781-784.

Kahneman, D. (1973). *Attention and effort.* Englewood Cliffs, N.J.: Prentice Hall.

Kaiser, J., & Lutzenberger, W. (2005). Human gamma-band activity: A window to cognitive processing. *Neuroreport*, 16, 207-211.

Kalcher, J., & Pfurtscheller, G. (1995). Discrimination between phase-locked and non-phase-locked event-related EEG activity. *Electroencephalography and Clincial Neurophysiology*, 94, 381-384.

Kalsbeek, J., & Ettema, J. (1963). Continuous recording of heart rate and the measurement of perceptual load. *Ergonomics*, 6, 306-307.

Kamijo, K., Nishihira, Y., Higashiura, T., & Kuroiwa, K. (2007). The interactive effect of excercise intensity and task difficulty on human cognitive processing. *International Journal of Psychophysiology*, 65, 114-121.

Kaminer, W. (2000). *Russendisko* [audio book]. Germany: BMG Wort.

Katayama, J., & Polich, J. (1998). Stimulus context determines P3a and P3b. *Psychophysiology*, 35, 23-33.

Kecklund, G., & Åkerstedt, T. (1993). Sleepiness in long distance truck driving: an ambulatory EEG study of night driving. *Ergonomics*, 36, 1007-1017.

Keil, A., Müller, M.M., Ray, W.J., Gruber,T., & Elbert, T. (1999). Human gamma band activity and perception of a gestalt. *The Journal of Neuroscience*, 19, 7152-7161.

Kinomura, S., Larsson, J., Gulyas, B., & Roland, P. E. (1996). Activation by attention of the human reticular formation and thalamic intralaminar nuclei. *Science*, 271, 512–515.

Klimesch, W. (1996). Memory processes, brain oscillations and EEG synchronization. *International Journal of Psychophysiology*, 24, 61-100.

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

Klimesch, W. (1997). EEG-alpha rhythms and memory processes. *International Journal of Psychophysiology*, 26, 319-340.

Klimesch, W., Doppelmayr, M., Röhm, D., Pöllhuber, D., & Stadler, W. (2000). Simultaneous desynchronization and synchronization of different alpha responses in the human electroencephalograph: a neglected paradox? *Neuroscience Letters*, 284, 97-100.

Klimesch, W., Doppelmayr, M., Russegger, H., & Pachinger, T. (1996). Theta band power in the human scalp EEG and the encoding of new information. *Neuroreport*, 7, 1235-1240.

Klimesch, W., Doppelmayr, M., Schwaiger, J., Auinger, P., & Winkler, T. (1999). 'Paradoxical' alpha synchronization in a memory task. *Cognitive Brain Research*, 7, 493-501.

Klimesch, W., Sauseng, P., & Hanslmayr, S. (2007). EEG alpha oscillations: the inhibition-timing hypothesis. *Brain Research Reviews*, 53, 63-88.

Klimesch, W., Schimke, H., Ladurner, G.S., & Pfurtscheller, G. (1990). Alpha frequency and memory performance. *Journal of Psychophysiology*, 4, 381-390.

Klimesch, W., Schimke, H., & Pfurtscheller, G. (1993). Alpha frequency, cognitive load and memory performance. *Brain Topography*, 5, 241-251.

Klinger, E., Gregoire, K.C., & Barta, S.G., (1973). Physiological correlates of mental activity: eye movements, alpha, and heart rate during imaging, suppression, concentration, search, and choice. *Psychophysiology*, 10, 471-477.

Kohlmorgen, J., Dornhege, G., Braun, M.L., Blankertz, B., Müller, K.-R., Curio, G., et al. (2007). Improving human performance in a real operating environment through real-time mental workload detection. In G. Dornhege, J. d. R. Milán, T. Hinterberger, D. McFarland & K.-R. Mueller (Eds.), *Toward Brain-Computer Interfacing* (pp. 409-422). Cambridge: MIT press.

Kok, A. (1997). Event-related-potential (ERP) reflections of mental resources: a review and synthesis. *Biological Psychology*, 45, 19-56.

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

Kramer, A.F. (1991). Physiological metrics of mental workload: A review of recent progress. In D.L. Damos (Ed.), *Multiple task performance* (pp. 279-328). London: Taylor & Francis.

Kramer, A., Sirevaag, E., & Hughes, P. (1988). Effects of foveal task load on visual-spatial attention: Event-related brain potentials and performance. *Psychophysiology*, 25, 512-531.

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

Krems, J.F., Keinath, A., Baumann, M., & Jahn, G. (2004). Die Okklusionsmethode: Ein einfaches und valides Verfahren zur Bewertung der visuellen Beanspruchung von Zweitaufgaben? In B. Schlag (Ed.), *Verkehrspsychologie* (pp. 335-349). Lengerich, Pabst.

Kubota, Y., Sato, W., Toichi, M., Murai, T., Okada, T., Hayashi, A. et al. (2001). Frontal midline theta rhythm is correlated with cardiac autonomic activities during the performance of an attention demanding meditation procedure. *Cognitive Brain Research*, 11, 281-287.

Lacey, J. (1967). Somatic response patterning and stress: some revisions of activation theory. In M.H. Appley & R. Trumbull (Eds.), *Psychological Stress: Issues in Research.* New York: Appleton Century Crofts.

Lal, S.K.L., & Craig, A. (2001a). A critical review of the psychophysiology of driver fatigue. *Biological psychology*, 55, 173-194.

Lal, S.K.L., & Craig, A. (2001b). Electroencephalography activity associated with driver fatigue implications for a fatigue countermeasure device. *Journal of Psychophysiology*, 15, 183-189.

Lal, S.K.L., Craig, A., Boord, P., Kirkup, L., & Nguyen, H. (2003). Development of an algorithm for an EEG-based driver fatigue countermeasure. *Journal of Saftey Research*, 34, 321-328.

Lamble, D., Kauranen, T., Laasko, M., & Summala, H. (1999).Cognitive load and detection thresholds in car following situations: safety implications for using mobile (cellular) telephones while driving. *Accident Analysis and Prevention*, 31, 617-623.

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

Lemke, M. (1982). Correlation between EEG and driver's actions during prolonged driving under montonous conditions. *Accident Analysis and Prevention*, 14, 7-17.

Levy, J., Pashler , H., & Boer, E. (2006). Central interference in driving. Is there any stopping the psychological refractory period? *Psychological Science*, 17, 228-235.

Lin, C.-T., Wu, R.C., Liang, S.-F., Chao, W.-H., Chen, Y.-J., & Jung, T.-P. (2005). EEG-based drowsiness estimation for safety driving using independent component analysis. *IEEE Transactions onCircuits and Systems-I: Regular Papers*, 52, 2726-2738.

Lins, O.G., Picton, T.W., Berg, P., & Scherg, M. (1993a). Ocular artifacts in EEG and event-related potentials I: scalp topography. *Brain Topography*, 6, 51-63.

Lins, O.G., Picton, T.W., Berg, P., & Scherg, M. (1993b). Ocular artifacts in recording EEGs and event-related potentials. II: Source dipoles and source components. *Brain Topography*, 6, 65-78.

Lissy, K., Cohen, J. Park, M., & Graham, J.D. (2000). Cellular phones and driving: weighing the risks and benefits. *Risk in Perspective*, 8, 1-6.

Lorist, M.M., Snel, J., Kok, A., & Mulder, G. (1994). The influence of caffeine on selective attention in well rested and fatigued subjects. *Psychophysiology*, 31, 525-534.

Luck, S.J. (2005). *An introduction to the event-related potential technique.* Cambridge, USA: The MIT Press.

Luck, S.J., & Hillyard, S.A. (1994). Electrophysiological correlates of feature analysis during visual search. *Psychophysiology*, 31, 291-308.

Luck, S.J., Girelli, M., McDermott, M.T., & Ford, M.A. (1997). Bridging the gap between monkey neurophysiology and human perception: An ambiguity resolution theory of visual selective attention. *Cognitive Psychology*, 33, 64-87.

Luczak (1987). Psychophysiologische Methoden zur Erfassung psychologischer Beanspruchungszustände. In U. Kleinbeck & J. Rutenfranz (Eds.), *Arbeitspsychologie. Enzyklopädie der Psychologie: Vol. D/III/1.* (pp. 185-259). Göttingen: Hogrefe.

Mangun, G. R., & Hillyard, S.A. (1990). Allocation of visual attention to spatial locations: Trade-off functions for event-related brain potentials and detection performance. *Perception and Psychophysics*, 47, 532-550.

Mangun, G. R., & Hillyard, S.A. (1997). Mechanisms and models of selective attention. In M.G.H. Coles & M.D. Rugg (Eds.), *Electrophysiology of mind. Event-related brain potentials and cognition* (pp. 1-26). Oxford: Oxford University Press.

Mangun, G.R. (1995). Neural mechanisms of visual selective attention. *Psychophysiology*, 32, 4-18.

Mann, C.A., Sterman, M.B., & Kaiser, D.A. (1996). Suppression of EEG rhythmic frequencies during somato-motor and visuo-motor behavior. *International Journal of Psychophysiology*, 23, 1-7.

Manzey, D. (1998). Psychophysiologie mentaler Beanspruchung. In F. Roesler (Ed.), Ergebnisse und Anwendungen der Psychophysiologie. *Enzyklopädie der Psychologie: Vol. C/I/5. Biologische Psychologie* (pp. 799-864). Göttingen: Hogrefe.

Mattes, S. (2003). The lane-change-task as a tool for driver distraction evaluation. In H. Strasser, K. Kluth, H. Rausch & H. Bubb (Eds.), *Quality of work and products in enterprises of the future* (pp. 57-60). Stuttgart: Ergonomia Verlag.

McCarthy, G., & Donchin, E. (1981). A metric for thought: a comparison of P300 latency and reaction time. *Science*, 211, 77-80.

McEvoy, S.P., Stevenson, M.R., McCartt, A.T., Woodwart, M., Haworth, C., Palamara, P., et al. (2005). Role of mobile phones in motor vehicle crashes resulting in hospital attendance: a case-crossover study. *BMJ*, doi:10.1136/bmj.38537.397512.55 (published 12 July 2005).

McKenna, P., Jefferies, L., Dobson, A., & Frude, N. (2004). The use of a cognitive battery to predict who will fail an on-road driving test. *British Journal of Clinical Psychology*, 43, 325-336.

McKnight, A.J., & McKnight A.S. (1993). The effect of c ellular phone use upon driver attention. *Accident Analysis and Prevention*, 25, 259-265.

Mecklinger, A., Kramer, A.F., & Strayer, D.I. (1992). Event related potentials and EEG components in a semantic memory search task. *Psychophysiology*, 29, 104-119.

Menon, V., Adleman, N. E., White, C. D., Glover, G. H., & Reiss, A. L. (2001). Error-related brain activation during a Go/NoGo response inhibition task. *Human Brain Mapping*, 12, 131–143.

Michon, J.A. (1985). A critical view of driver behavior models: What do we know, what should we do? In L. Evans & R. Schwing (Eds.), *Human Behavior and Traffic Safety* (pp. 485-520). New York: Plenum Press.

Michon, J.A. (1993). *Generic intelligent driver support. A comprehensive report on GIDS.* London: Taylor & Francis.

Miller, J.C. (1995). Batch processing of 10000 h of truck driver EEG data. *Biological Psychology*, 40, 209-222.

Muckler, F.A., & Seven, S.A. (1992). Selecting performance measures: 'objective' versus 'subjective' measurement. *Human Factors*, 34, 441-455.

Mulder, G. (1980). *The heart of mental effort*. Groningen: University of Groningen.

Mulder, G., & Mulder, L.J.M. (1981). Information processing and cardiovascular control. *Psychophysiology*, 14, 392-402.

Mulder, L.J.M., & Mulder, G. (1987). Cardiovascular reactivity and mental workload. In O. Romplemann & R.I. Kitney (Eds.), *The beat-by-beat investigation of cardiovascular function* (pp. 216-253). Oxford: Oxford University Press.

Näätänen, R. (1982). Processing negativity: an evoked potential reflection of selective attention. *Psychological Bulletin*, 92, 605-640.

Näätänen, R. (1990). The role of attention in auditory information processing as revealed by event-related potentials and other brain measures of cognitive function. *Behavioral and Brain Sciences*, 13, 201-288.

Näätänen, R., & Picton, T. (1987). The N1 wave of the human electric and magnetic response to sound: A review and an analysis of the component structure. *Psychophysiology*, 24, 375-425.

Nachreiner, F., & Schultetus, W. (2002). Normung im Bereich der psychischen Belastung – die Normen der Reihe DIN EN ISO 10075. *DIN-Mitteilung*, 81, 519-524.

Naumann, E., Maier, S., Diedrich, O., Becker, G., & Bartussek, D. (1997). Structural, semantic, and emotion-focused processing of neutral and negative nouns: Event-related potential correlates. *Journal of Psychophysiology*, 11, 158-172.

Nordbakke, S., & Sagberg, F. (2007). Sleepy at the wheel: knowledge, symptoms and behaviour among car drivers. *Transportation Research Part F*, 10, 1-10.

Norman, D.A., & Shallice, T. (1986). Attention to action: willed and automatic control of behavior. In R. Davidson, G. Schwartz & D. Shapiro (Eds.), *Consciousness and self regulation: Advances in research and theory* (pp. 1-18). New York: Plenum.

Nunes, L., & Recarte, M.A. (2002). Cognitive demands of hands-free-phone conversation while driving. *Transportation research part F*, 5, 133-144.

O'Donnell, R.D., & Eggemeier, F.T. (1986). Workload assessment methodology. In K.R. Boff, L. Kaufman, & J.P. Thomas (Eds.), *Handbook of perception and human performance: Vol. II. Cognitive processes and performance* (pp. 42/1 - 42/49). New York: Wiley.

Oldfield, R.C. (1971). The assessment and analysis of handedness. The Edinburgh inventory. *Neuropsychologia*, 9, 97-113.

Palva, J.M, Palva, S., & Kaila, K. (2005). Phase synchrony among neuronal oscillations in the human cortex. *The Journal of Neuroscience*, 25, 3962-3972.

Palva, S., & Palva, J.M. (2007). New vistas for α-frequency band oscillations. *Trends in Neurosciences*, 30, 150-158.

Parasuraman, R. (1978). Auditory evoked potentials and divided attention. *Psychophysiology*, 15, 460-465.

Parasuraman, R. (1985). Event-related brain potentials and intermodal divided attention. *Proceedings of the Human Factors Society*, 29, 971-975.

Parasuraman, R. (1990). Event-related brain potentials and human factors research. In J.W. Rohrbaugh, R. Parasuraman & R.Jr. Johnson (Eds.), *Event-related brain potentials: Basic issues and applications* (pp. 279-300). Oxford: Oxford University Press.

Pashler, H., & Johnston, J.C. (1998). Attentional limitations in dual-task performance. In H. Pashler (Ed.), *Attention* (pp. 155-189). East Sussex: Psychology Press Ltd.

Patel, A.H., & Azzam, P.N. (2005). Characterization of N200 and P300: Selected studies of the event-related potential. *International Journal of Medical Sciences*, 2, 147-154.

Patten, C.J.D., Kirchner, A., Östlund, J., & Nilsson, L. (2004). Using mobile telephones: cognitive workload and attention resource allocation. *Accident Analysis and Prevention*, 36, 341-350.

Pfurtscheller, G. (2001). Functional brain imaging based on ERD/ERS. *Vision Research*, 41, 1257-1260.

Pfurtscheller, G. (2003). Induced oscillations in the alpha band: functional meaning. *Epilepsia*, 44, 2-8.

Pfurtscheller, G., Stancák, A. Jr., & Neuper, C. (1996). Event-related synchronization (ERS) in the alpha band – an electrophysiological correlate of cortical idling: A review. *International Journal of Psychophysiology*, 24, 39-46.

Picton, T.W., Alain, C., Otten, L., Ritter, W., & Achim, A. (2000). Mismatch negativity: Different water in the same river. *Audiology and Neuro-Otology*, 111-139.

Piechulla, W. (2006). Messung und Kontrolle mentaler Fahrerbeanspruchung. In U.W. Ebner-Priemer (Ed.), *Psychophysiologie in Labor und Feld. Ambulantes psychophysiologisches Monitoring – neue Perspektiven und Anwendungen* (pp. 29-59). Frankfurt a.M.: Peter Lang GmbH, Europäischer Verlag der Wissenschaften.

Piechulla, W., Mayser, C., Gehrke, H., & König, W. (2003). Reducing drivers' mental workload by means of an adaptive man-machine interface. *Transportation Research Part F*, 6, 233-248.

Polich T.W., & Criado, J.R. (2006). Neuropsychology and neuropharmacology of P3a and P3b. *International Journal of Psychophysiology*, 60, 172-185.

Pope, A.T., Bogart, E.H., & Bartolome, D.S. (1995). Biocybernetic system evaluates indices of operator engagement in automated task. *Biological Psychology,* 40, 187– 195.

Posner, M.L. (1980). Orienting of attention. *Quarterly Journal of Experimental Psychology*, 32, 3-25.

Potts, G.F., Liotti, M., Tucker, D.M., & Posner, M.I. (1996). Frontal and inferior temporal cortical activity in visual target detection: Evidence from high spatially sampled event-related potentials. *Brain Topography*, 9, 3-14.

Ranney, T.A. (1994). Models of driving behavior: A review of their evolution. *Accident Analysis and Prevention*, 26, 733-750.

Rantanen, E.M., & Goldberg, J.H. (1999). The effect of mental workload on the visual field size and shape. *Ergonomics*, 42, 816-834.

Rasmussen, J. (1983). Skills, rules, and knowledge: signals, signs, and symbols, and other distinctions in human performance models. *IEEE Transactions on Systems, Man, and Cybernetics*, SMC-13, 257-266.

Rauch, N., Totzke, I., & Krüger, H.-P. (2004). *Workload-abhängiger Warn- und Informations-Manager, Meilenstein I/2003, Literaturanalyse.* Veitshöchheim: Institute for Traffic Sciences.

Ravden, D., & Polich, J. (1998). Habituation of P300 from visual stimuli. *International Journal of Psychophysiology*, 30, 359-365.

Ray, W.J., & Cole, H.W. (1985a). EEG alpha activity reflects attentional demands, and beta activity reflects emotional and cognitive processes. *Science*, 228, 750-752.

Ray, W.J., & Cole, H.W. (1985b). EEG activity during cognitive processing: influence of attentional factors. *International Journal of Psychophysiology*, 3, 43-48.

Razmjou, S., & Kjellberg, A. (1992). Sustained attention and serial responding in heat: Mental effort in the control of performance. *Aviation, Space and Environmental Medicine*, 63, 594-601.

Recarte, M.A.. & Nunes, L. (2002). Mental load and loss of control over speed in real driving. Towards a theory of attentional speed control. *Transportation Research Part F*, 111-122.

Recarte, M.A., & Nunes, L.M. (2000). Effects of verbal and spatial-imagery task on eye fixations while driving. *Journal of Experimental Psychology: Applied*, 6, 31-43.

Recarte, M.A.. & Nunes, L.M. (2003). Mental workload while driving: effects on visual search, discrimination, and decision making. *Journal of Experimental Psychology: Applied*, 9, 119-137.

Recarte, M.A. Nunes, L.M., & Conchillo, A. (n.d.). Attention and eye-movements while driving: Effects of verbal versus spatial-imagery and comprehension versus response production tasks. Manuscript submitted for publication.

Redelmeier, D.A., & Tibshirani, R.J. (1997). Association between cellular-telephone calls and motor vehicle collisions. *The New England Journal of Medicine*, 336, 453-458.

Reid, G.B., & Nygren, T.E. (1988). The subjective workload assessment technique: A scaling procedure for measuring mental workload. In P.A. Hancock & N. Meshkati (Eds.), *Human mental workload* (pp. 185-213). Amsterdam: North Holland.

Ribback, S. (2003). *Psychophysiologische Untersuchung mentaler Beanspruchung in simulierten Mensch-Maschine-Interaktionen.* Doctoral dissertation, University of Potsdam, Germany, 2004 (URN No. urn:nbn:de:kobv:517-0000833). Retrieved February 21, 2008, from http://deposit.d-nb.de/cgi-bin/dokserv?idn=97002021x

Rippon, G. (2006). Electroencephalography. In C. Senior, T. Russell & M.S. Gazzaniga (Eds.), *Methods in mind* (pp. 237-262). Cambridge: The MIT Press.

Rohmert, W. (1984). Das Belastungs-Beanspruchungs-Konzept. *Zeitschrift für Arbeitswissenschaft*, 38, 193-200.

Roscoe, A.H. (1992). Assessing pilot workload. Why measure heart rate, HRV and respiration? *Biological Psychology*, 34, 259-287.

Rugg, M.D., & Coles, M.G.H. (1997). *Electrophysiology of mind. Event-related brain potentials and cognition.* Oxford: Oxford University Press.

Rugg, M.D., & Dickens, A.M.J. (1982). Dissociation of alpha and theta activity as a function of verbal and visuospatial tasks. *Electroencephalography and Clinical Neurophysiology*, 53, 201-207.

Ryu, K., & Myung, R. (2005). Evaluation of mental workload with a combined measure based on physiological indices during a dual task of tracking and mental arithmetic. *International Journal of Industrial Ergonomics*, 35, 991-1009.

Sams, M., Alho, K., & Näätänen, R. (1983). Sequential effects on the ERP in discriminating two stimuli. *Biological Psychology*, 17, 41-58.

Sanders, A.F. (1983). Towards a model of stress and human performance. *Acta Psychologica*, 53, 61-97.

Sauseng, P., Klimesch, W., Doppelmayr, M., Pecherstorfer, T., Freunberger, R., & Hanslmayr, S. (2005). EEG alpha synchronization and functional coupling during top-down processing in a working memory task. *Human Brain Mapping*, 26, 148-155.

Schack, B., Klimesch, W., & Sauseng, P. (2005). Phase synchronization between theta and upper alpha oscillations in a working memory task. *International Journal of Psychophysiology*, 57, 105-114.

Schandry, R. (2006). *Biologische Psychologie.* Weinheim: Beltz Verlag

Schier, M.A. (2000). Changes in EEG alpha power during simulated driving: a demonstration. *International Journal of Psychophysiology*, 155-162.

Schimke, H., Klimesch, W., & Pfurtscheller, G. (1990). Ereignisbezogene Desynchronization und die Wahl des Alpha-Frequenzbandes zur Quantifizierung kortikaler Prä- und Poststimulus-Aktivierung. *Zeitschrift für Elektroenzephalographie, Elektromyographie und verwandte Gebiete*, 20, 219-225.

Schlegel, R.E., Gilliland, K., & Schlegel, B. (1986). Development of the criterion task set performance data base. *Proceedings of the 30th annual meeting of the Human Factors Society* (pp. 58-62). Santa Monica, CA: Human Factors Society.

Schmidt, E. (2006). *Elektrophysiologische Parameter der Fahrerbeanspruchung im Feldexperiment.* Unpublished Diplomarbeit, Heinrich-Heine University Düsseldorf, Germany.

Schneider, W., & Shiffrin, R.M. (1977). Controlled and automatic human information processing: I. Detection search and attention. *Psychological Review*, 84, 1-54.

Schulz, A.P. (2007). *Driving without awareness. Folgen herabgesetzter Aufmerksamkeit im Straßenverkehr.* Saarbrücken: VDM Verlag Dr. Müller.

Schumacher, M. (2001). *Vergleich verschiedener subjektiver Maße der Beanspruchung bei Realfahrten im PKW.* Unpublished Diplomarbeit, Bayerische Julius-Maximilians-Universität Würzburg, Germany.

Schupp, H.T., Lutzenberger, W., Birbaumer, N., Miltner, W., & Braun, C. (1994). Neurophysiological differences between perception and imagery. *Cognitive Brain Research*, 2, 77-86.

Sederberg, P. B., Schulze-Bonhage, A., Madsen, J.R., Bromfield, E.B., McCarthy, D.C., Brandt, A., et al. (2007). Hippocampal and neocortical gamma oscillations predict memory formation in humans. *Cerebral Cortex*, 17, 1190-1196.

Shallice, T., & Burgess, P. (1993). Supervisory control of action and thought selection. In A. Baddeley & L. Weiskrantz (Eds.), *Attention: selection, awareness, and control. A tribute to Donald Broadbent* (pp. 171-187). Oxford: Clarendon Press.

Shiffrin, R.M., & Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending and a general theory. *Psychological Review*, 84, 127-189.

Simons, R.F., Graham, F.K., Miles, M.A., & Chen, X. (2001). On the relationship of the P3a and the novelty-P3. *Biological Psychology*, 56, 207-218.

Simson, R., Vaughan, H.G., & Ritter, W. (1977).The scalp topography of potentials in auditory and visual discrimination tasks. *Electroencephalography and Clinical Neurophysiology*, 42, 528-535.

Smith, M.E., Halgren, E., Sokolik, M., Baudena, P., Musolino, A., Liegeois-Chauvel, C., et al. (1990). The intracranial topography of the P3 event-related potential elicited during auditory oddball. *Electroencephalography and Clinical Neurophysiology*, 76, 235-248.

Snodgrass, J.G., & Corwin, J. (1988). Pragmatics of measuring recognition memory: Applications to dementia and amnesia. *Journal of Experimental Psychology: General*, 117, 34-50.

Sterman, M.B. (1981). EEG-feedback: Physiological behavior modification. *Neuroscience and Biobehavioral Review*, 5, 405-412.

Strayer, D. L., & Drews, F.A. (2007). Cell-phone-induced driver distraction. *Current Directions in Psychological Science*, 16, 128-131.

Strayer, D.L., & Johnston, W.A. (2001). Driven to distraction: dual-task studies of simulated driving and conversing on a cellular phone. *Psychological Science*, 12, 462-466.

Strayer, D.L., Drews, F.A., & Crouch, D.J. (2006). A Comparison of the Cell Phone Driver and the Drunk Driver. *Human Factors*, 48, 381-391.

Strayer, D.L., Drews, F.A., & Johnston, W.A. (2003). Cell phone-induced failures of visual attention during simulated driving. *Journal of Experimental Psychology: Applied*, 9, 23-32.

Sussmann, E.D., Bishop, H., Madnick, B., & Walter, W. (1985). Driver inattention and highway safety. *Transportation Research Record*, 1047, 40-48.

Sutton, S., Braren, M., Zubin, J., & John, E.R. (1965). Evoked-potential correlates of stimulus uncertainty. *Science*, 150, 1187-1188.

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, 654-662.

Tejero, P., & Chóliz, M. (2002). Driving on the motorway: the effect of alternating speed on driver's activation level and mental effort. *Ergonomics*, 45, 605-618.

Tsang, P.S., & Vidulich, M. A. (2006). Mental workload and situational awareness. In G. Salvendy (Ed.), *Handbook of human factors and ergonomics* (pp. 243-268). Hoboken: John Wiley and sons.

Tsang, P.S., & Wilson, G. (1997). Mental workload. In G. Salvendy (Ed.), *Handbook of human factors and ergonomics* (2nd ed., pp. 417-449). New York: Wiley.

Underwood, G. (2002). The application of psychology to driving behaviour. *Applied Cognitive Psychology*, 16, 359-366.

Verleger, R. (1988). Event-related potentials and cognition: A critique of the context updating hypothesis and an alternative interpretation of P3. *Behavioral and Brain Sciences*, 11, 343-427.

Verwey, W.B. (2000). On-line driver workload estimation. Effects of road stuation and age on secondary task measures. *Ergonomics*, 43, 187-209.

Verwey, W.B., & Janssen, W.H. (1988). *Route following and driving performance with in-car route guidance systems.* IZF-Report C-14. Soesterberg, NL: Institute for Perception TNO.

Volvo (2004, April 22). Das Sicherheitssystem IDIS. *Motor-Talk.* Retrieved August 22, 2007, from http://www.motor-talk.de/t296317/f93/s/thread.html.

Von Stein, A., & Sarnthein, J. (2000). Different frequencies for different scales of cortical integration: from local gamma to long range alpha/theta synchronization. *International Journal of Psychophysiology*, 38, 301-313.

Walter, H., Vetter, S.C., Grothe, J., Wunderlich, A.P., Hahn, S., & Spitzer, M. (2001). The neural correlates of driving. *Neuroreport*, 12, 1763-1767.

Ward, L.M. (2003). Synchronous neural oscillations and cognitive processes. *Trends in Cognitive Sciences*, 7, 553-559.

Welford, A.T. (1952). The “psychological refractory period” and the timing of high speed performance – a review and a theory. *British Journal of Psychology*, 43, 2-19.

Wester, A.E., Böcker, K.B.E., Volkerts, E.R., Vester, J.C., & Kenemans, J.L. (2008). Event-related potentials and secondary task performance during simulated driving. *Accident Analysis and Prevention*, 40, 1-7.

Wickens, C.D. (1984). Processing resources of attention. In R. Parasuraman & D.R. Davies (Eds.), *Varieties of attention* (pp. 63-102). Orlando, FL: Academic Press.

Wickens, C.D. (1992). *Engineering psychology and human performance.* New York: HarperCollins.

Wickens, C.D., & Hollands, J.G. (2000). *Engineering psychology and human performance.* London: Prentice-Hall.

Wickens, C., Kramer, A., Vanasse, L., & Donchin, E. (1983). Performance of concurrent tasks: A psychophysiological analysis of the reciprocity of information-processing resources. *Science*, 221, 1080-1082.

Wickens, C.D., Lee, J.D., Liu, Y., & Becker, S. E. G. (2004). *An introduction to Human Factors Engineering.* Upper Saddle River: Pearson Education.

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

Wijesuriya, N., Tran, Y., & Craig, A. (2007). The psychophysiological deteminants of fatigue. *International Journal of Psychophysiology*, 63, 77-86.

Willmann, S. (2003). *Stimulator* [computer software]. Stuttgart: DaimlerChrysler AG

Wilson, G., Fullenkamp, P., & Davis, I. (1994). Evoked potential, cardiac, blink, and respiration measures of pilot workload in air-to-ground missions. *Aviation, Space, and Environmental Medicine*, 65, 100-105.

Wilson, G.F., & Fisher, F. (1995). Cognitive task classification based upon topographic EEG data. *Biological Psychology*, 40, 239-250.

Wilson, G.F. (1991). Psychophysiological assessment of mental workload. In D.L. Damos (Ed.), *Multiple task performance* (pp. 279-328). London: Taylor & Francis.

Wilson, G.F. (2002). Flight Psychophysiology. *The International Journal of Aviation Psychology*, 12, 1-2.

Wilson, J. Fang, M., Wiggins, S., & Cooper, P. (2003). Collision and violation involvement of drivers who use cellular telephones. *Traffic Injury Prevention*, 4, 45-52.

Wiltschko, T. (2004). *Sichere Information durch infrastrukturgestützte Fahrerassistenzsysteme zur Steigerung der Verkehrssicherheit an Straßenknotenpunkten.* Fortschritt-Berichte VDI, 12. Duesseldorf: VDI Verlag.

Winn, P. (2001). *Dictionary of Biological Psychology*. London: Routledge.

Woldorff, M.G., Gallen, C.C., Hampson, S.A., Hillyard, S.A., Pantev, C., Sobel, D., et al. (1993). Modulation of early sensory processing in human auditory cortex during auditory selective attention. *Proceedings of the National Academy of Science*, 90, 8722-8726.

Woldorff, M.G., Hackley, S.A., & Hillyard, S.A. (1991). The effects of channel-selective attention on the mismatch negativity wave elicited by deviant tones. *Psychophysiology*, 28, 30-42.

Worden, M.S., Foxe, J.J., Wang, N., & Simpson, G.V. (2000). Anticipatory biasing of visuospatial attention indexed by retinotopically specific alpha-band electroencephalography increases over occipital cortex. *Journal of Neuroscience*, 20, 1-6.

Yamaguchi, S., & Knight, R.T. (1991). Anterior and posterior association cortex contributions to the somatosensory P300. *Journal of Neuroscience*, 11, 2039-2054.

Yeh, Y.-Y., & Wickens, C.D. (1988). The dissociation of subjective measures of mental workload and performance. *Human Factors*, 30, 111-120.

Zeier, H., & Bättig, K. (1977). Psychovegetative Belastung und Aufmerksamkeitsspannung von Fahrzeuglenkern auf Autobahnabschnitten mit und ohne Geschwindigkeitsbegrenzung. *Zeitschrift für Verkehrssicherheit*, 23, 11-15.

Zenker, F., & Barajas,.J.J. (1999). Auditory P300 development from an active, passive and single-tone paradigms. *International Journal of Psychophysiology*, 33, 99-111.

Zijlstra, F.R.H., & van Doorn, L. (1985). *The construction of a scale to measure perceived effort.* Delft, The Netherlands: Department of Philosophy and Social Sciences, Delft University of Technology.

Appendix A: Individual EEG Power Spectra of Three Experiments

In the following, the relative EEG power spectra for each subject in each experiment are displayed to illustrate the individual and task-specific alpha power adjustment procedure. Please note that for each subject the spectrum at the electrode showing maximal alpha power was selected. The black vertical lines indicate the 2 Hz range around the individual frequency peak in the alpha band which was extracted for the statistical power analysis. Please be also aware that spectra show different scales at the y-axis.

Figure 52 shows the individual alpha power peak in the baseline recording with eyes closed in laboratory Experiment 1. Some subjects (e.g. s03, s13) show very high power amplitudes (> 5 $\mu V^2/Hz^2$) while others barely show any alpha (e.g. s22 < 1 $\mu V^2/Hz^2$).

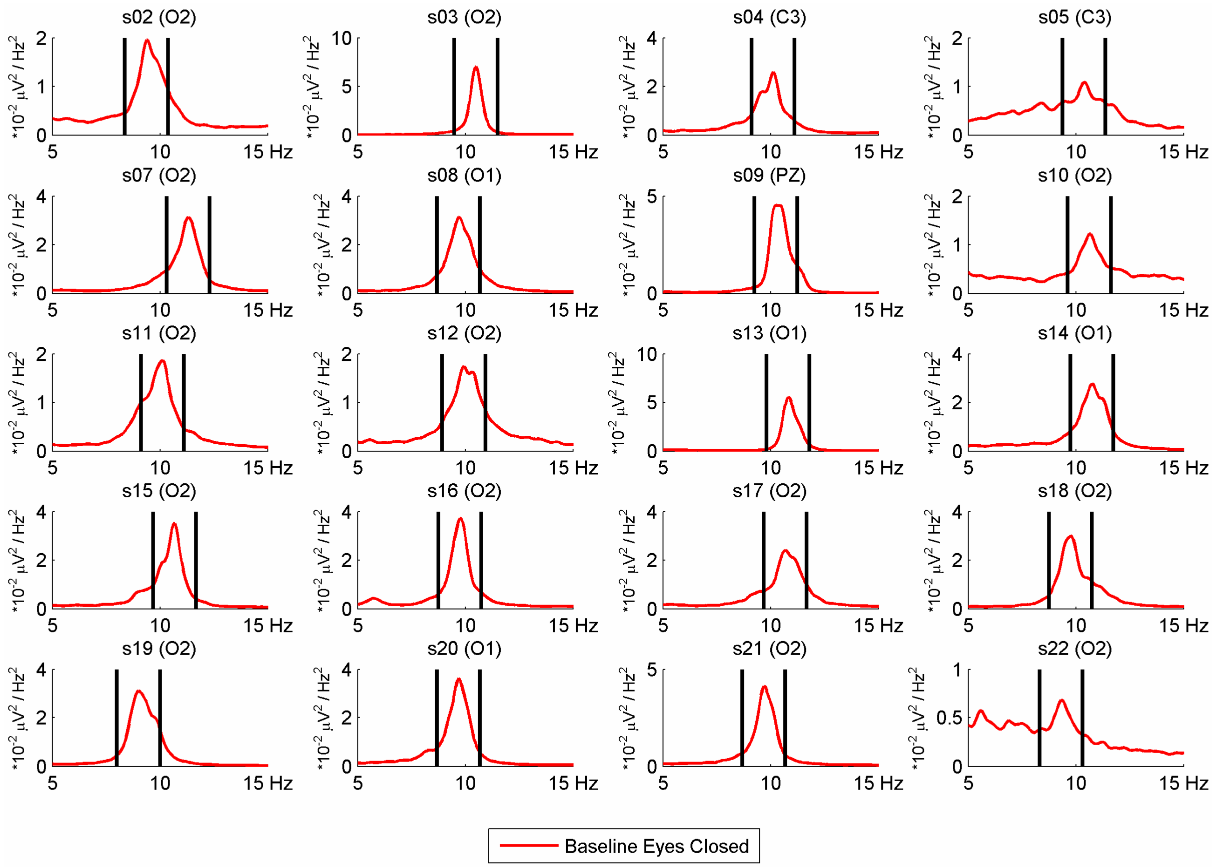

Figure 52. All 20 subjects' relative power spectra for the baseline recording with eyes closed in Experiment 1.

Figure 53 displays the individual power spectra in the condition in which subjects performed the cognitively demanding task without parallel driving. The majority of subjects show an increase of alpha power during the more demanding combined story listening task with word detection as opposed to the simple tone detection. A similar pattern of results can be observed in *Figure 54*, when subjects performed the two tasks under driving conditions, although the general power level is decreased in many subjects.

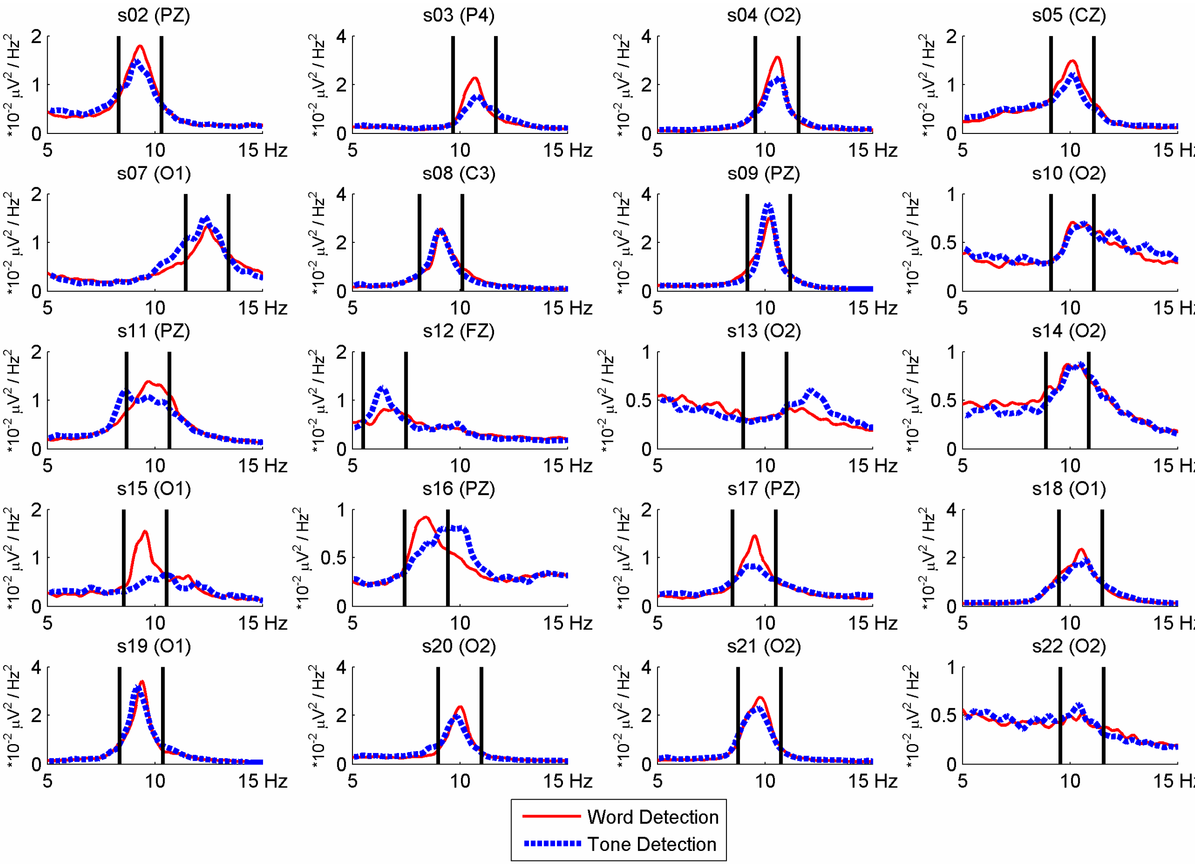

Figure 53. All 20 subjects' relative power spectra for the combined story listening task with word detection compared to the simple tone detection in Experiment 1.

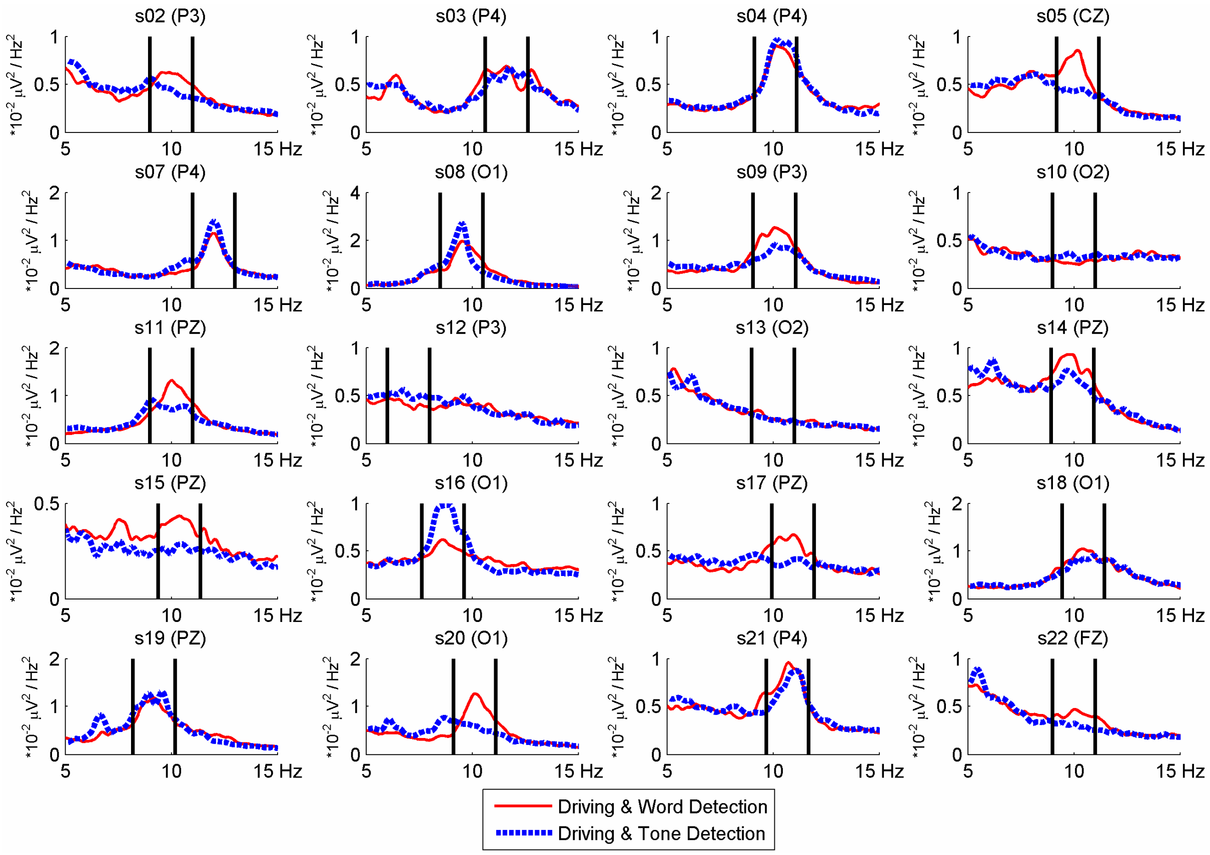

Figure 54. All 20 subjects' relative power spectra combined story listening task with word detection compared to the simple tone detection when simultaneously performing a Lane-Change Task in Experiment 1.

Figure 55 shows the individual alpha power peak in the baseline recording with eyes closed in Experiment 2. Again, some subjects (e.g. s02, s06) show very high power amplitudes (> 4.5 $\mu V^2/Hz^2$) while others barely show any alpha (e.g. s17 < 1 $\mu V^2/Hz^2$).

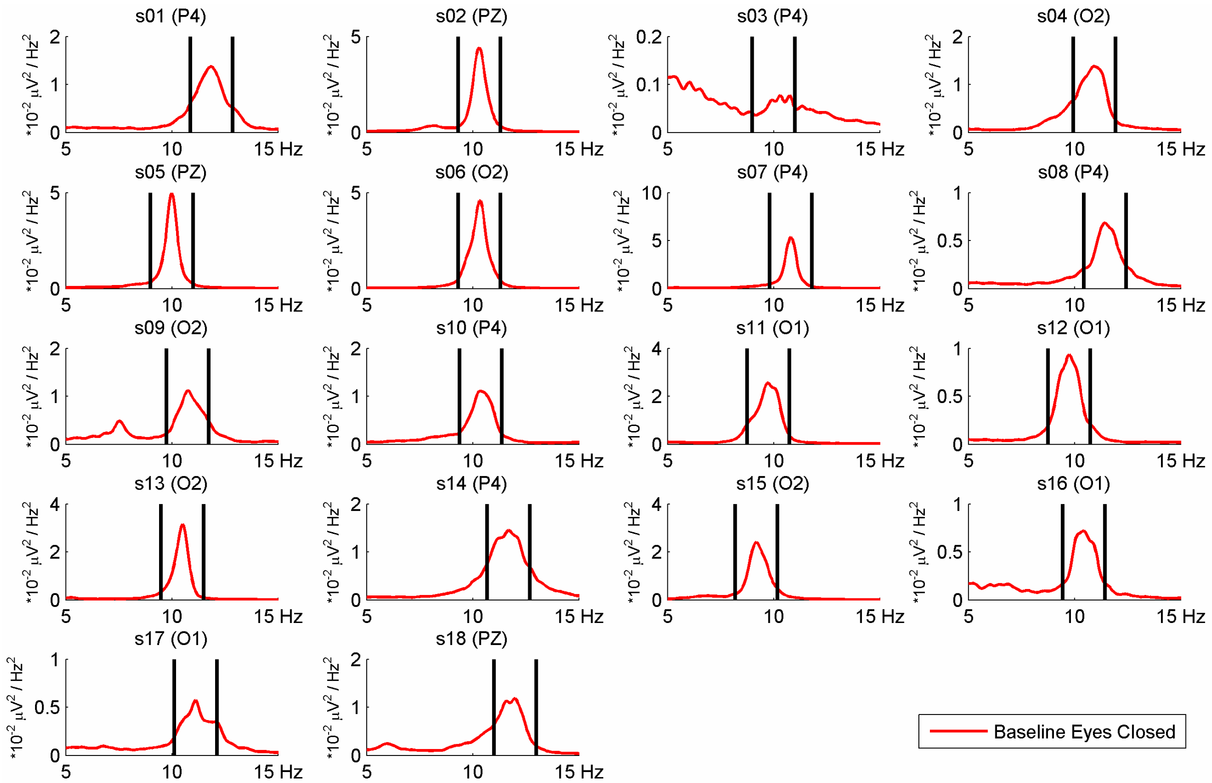

Figure 55. All 18 subjects' relative power spectra for the baseline recording with eyes closed in Experiment 2.

Figure 56 and *Figure 57* present the individual power spectra in the conditions in which subjects performed cognitively demanding tasks in parallel to driving in real road-traffic. The majority of subjects show an increase of alpha power during the more demanding cognitive tasks as opposed to driving sections without additional task.

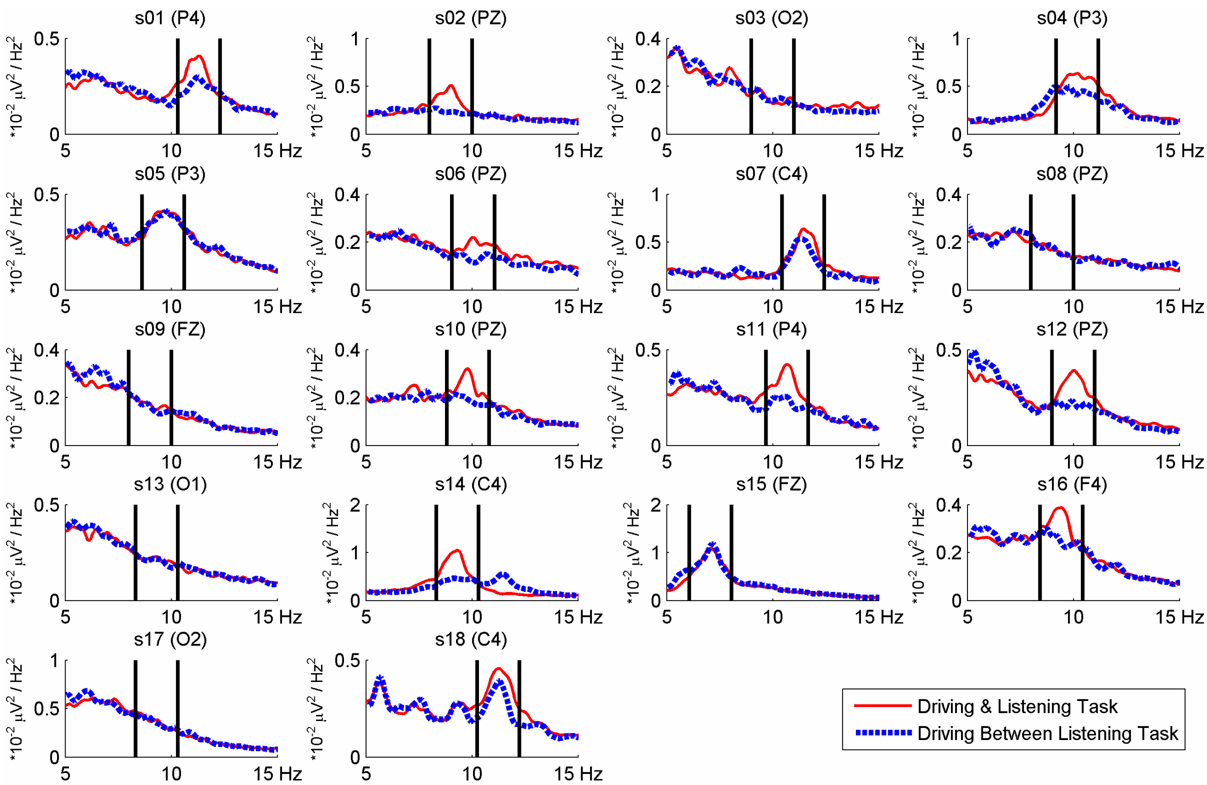

Figure 56. All 18 subjects' relative power spectra for the story listening task compared to driving without additional task in Experiment 2.

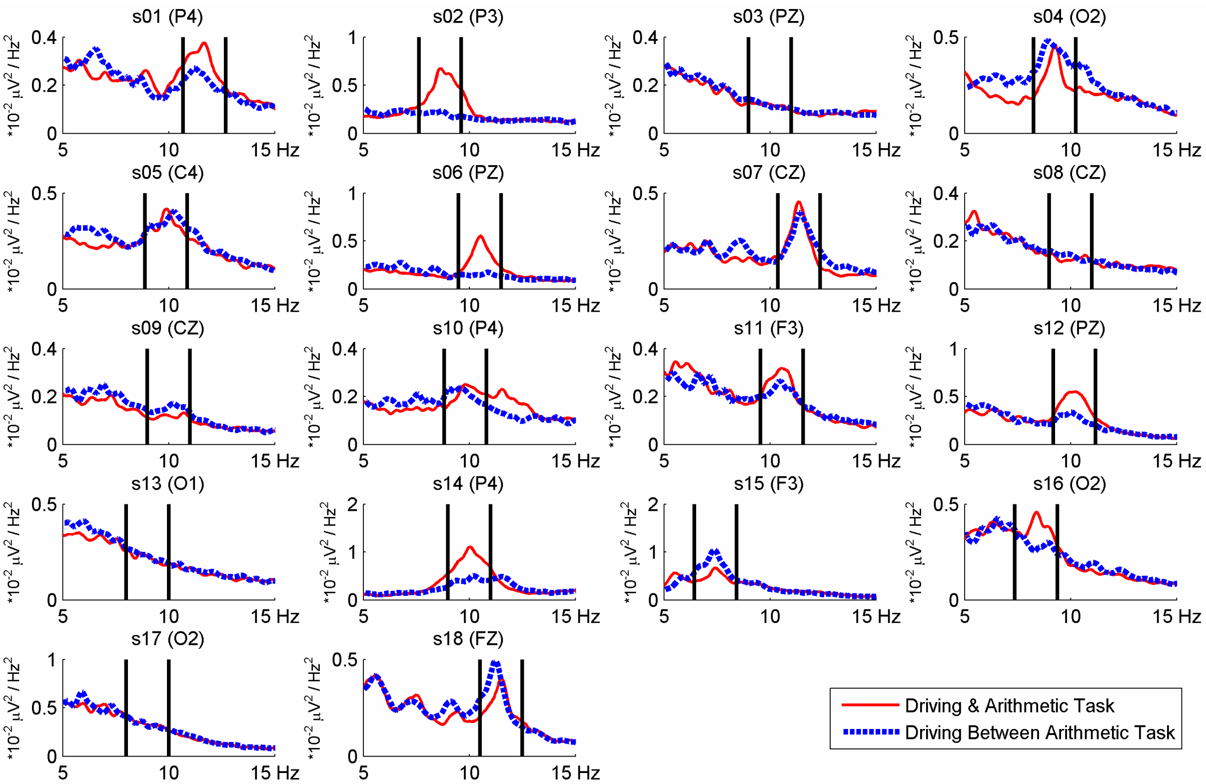

Figure 57. All 18 subjects' relative power spectra for the mental arithmetic task compared to driving without additional task in Experiment 2.

Figure 58 shows the individual alpha power peak in the baseline recording with eyes closed and in the two baseline driving conditions in Experiment 3. Again, during rest some subjects (e.g. s10, s20) show very high power amplitudes (> 4.5 µV²/Hz²) while others barely show any alpha (e.g. s62 < 2 µV²/Hz²). The qualitative differences in amplitude and frequency between the different baseline conditions in some subjects are clearly visible in this figure (e.g. s36, s48).

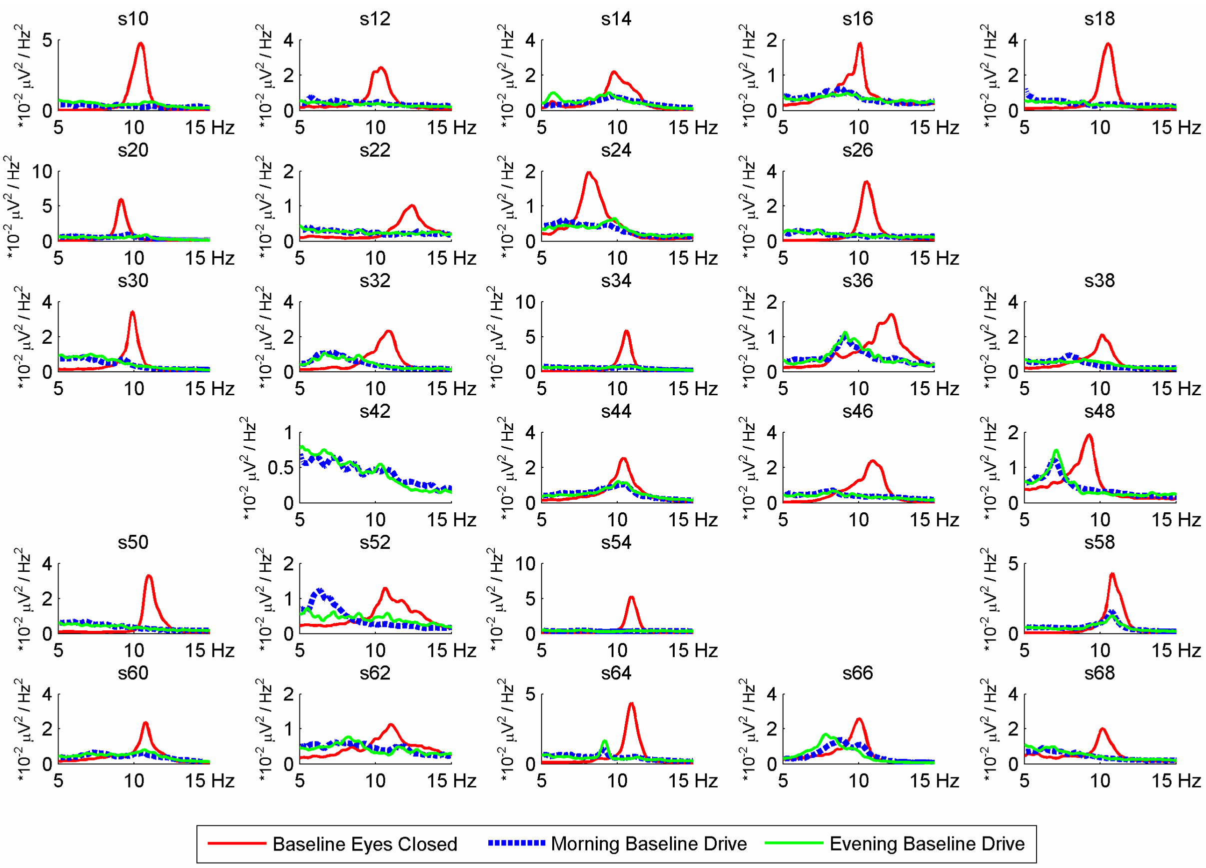

Figure 58. All 27 drivers' relative power spectra for the baseline recording with eyes closed and the two driving baselines at the beginning and at the end of Experiment 3.

Figure 59 and *Figure 60* show the individual power spectra in the conditions in which subjects performed the combined story listening task with word detection in parallel to driving on dirt roads of Experiment 3. The majority of subjects show an increase of alpha power during the more demanding cognitive tasks as opposed to driving sections without additional task.

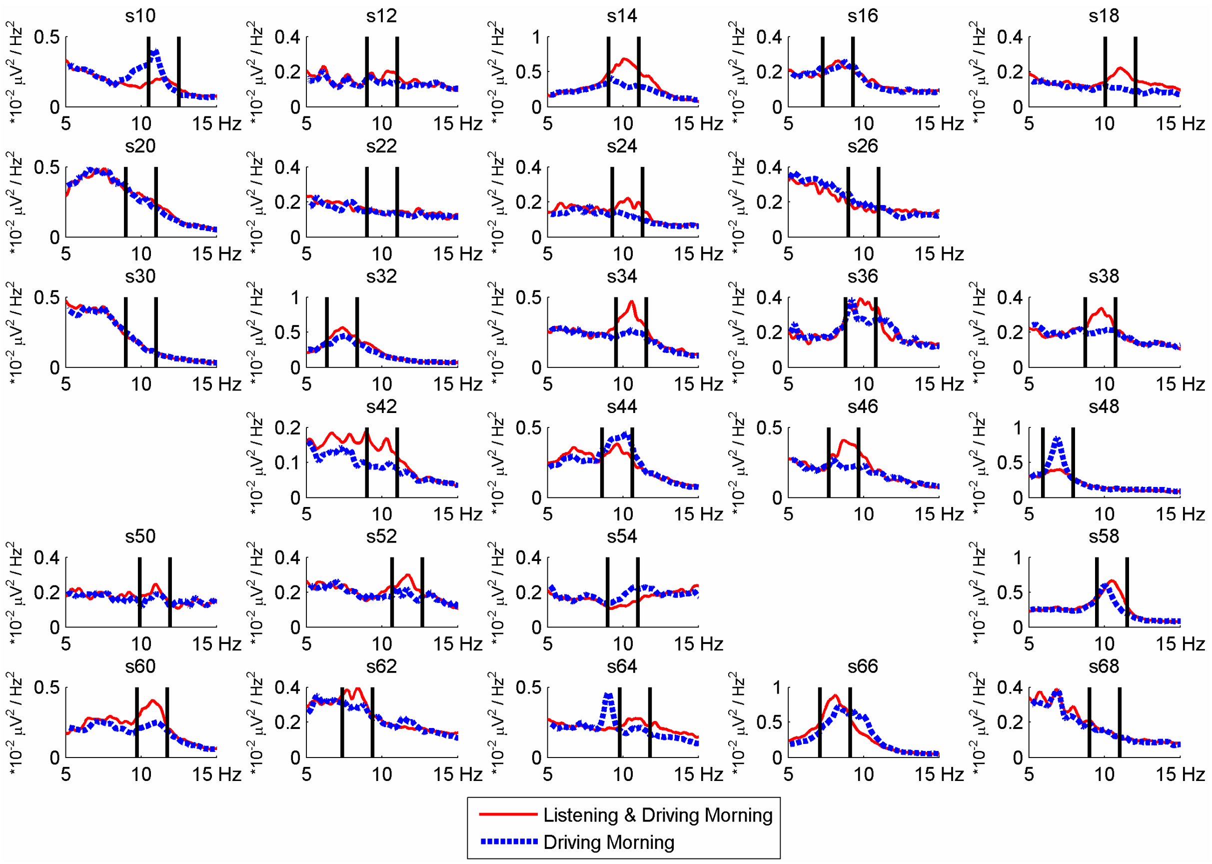

Figure 59. All 27 subjects' relative power spectra for the combined story listening task with word detection compared to driving without additional task in the morning session of Experiment 3.

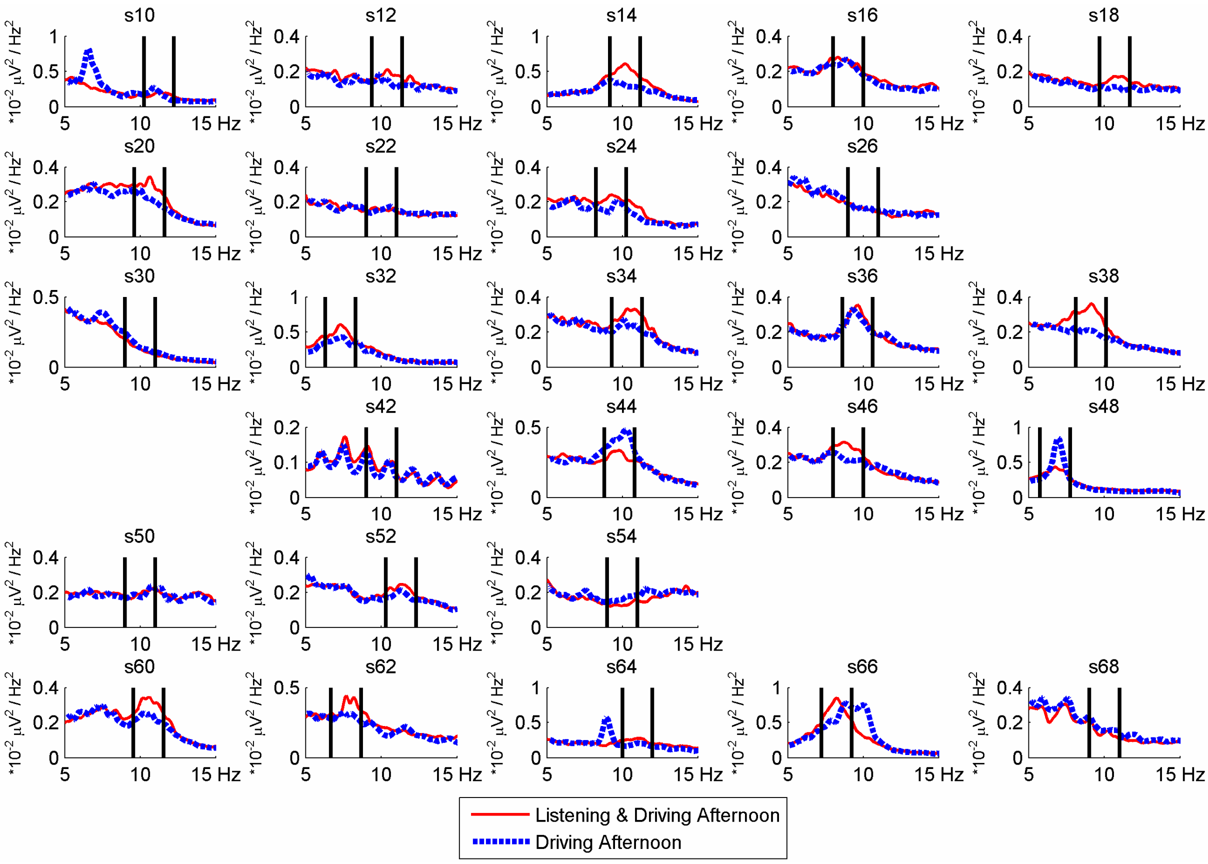

Figure 60. All 26 subjects' relative power spectra for the combined story listening task with word detection compared to driving without additional task in the afternoon session of in Experiment 3.

Appendix B: Standard EEG Power Spectra in Experiment 1

Alpha power effects in Experiment 1 were also significant when performing a grand-average in the standard alpha band (9-11 Hz) instead of an individual and task-specific alpha peak adjustment. *Figure 61* shows the grand-average over 20 subjects and 11 electrodes in the baseline condition with eyes closed in Experiment 1. A peak in the alpha band around 10 Hz which is strongest over occipital electrodes can be observed. The peak is equally distributed over both electrode hemispheres.

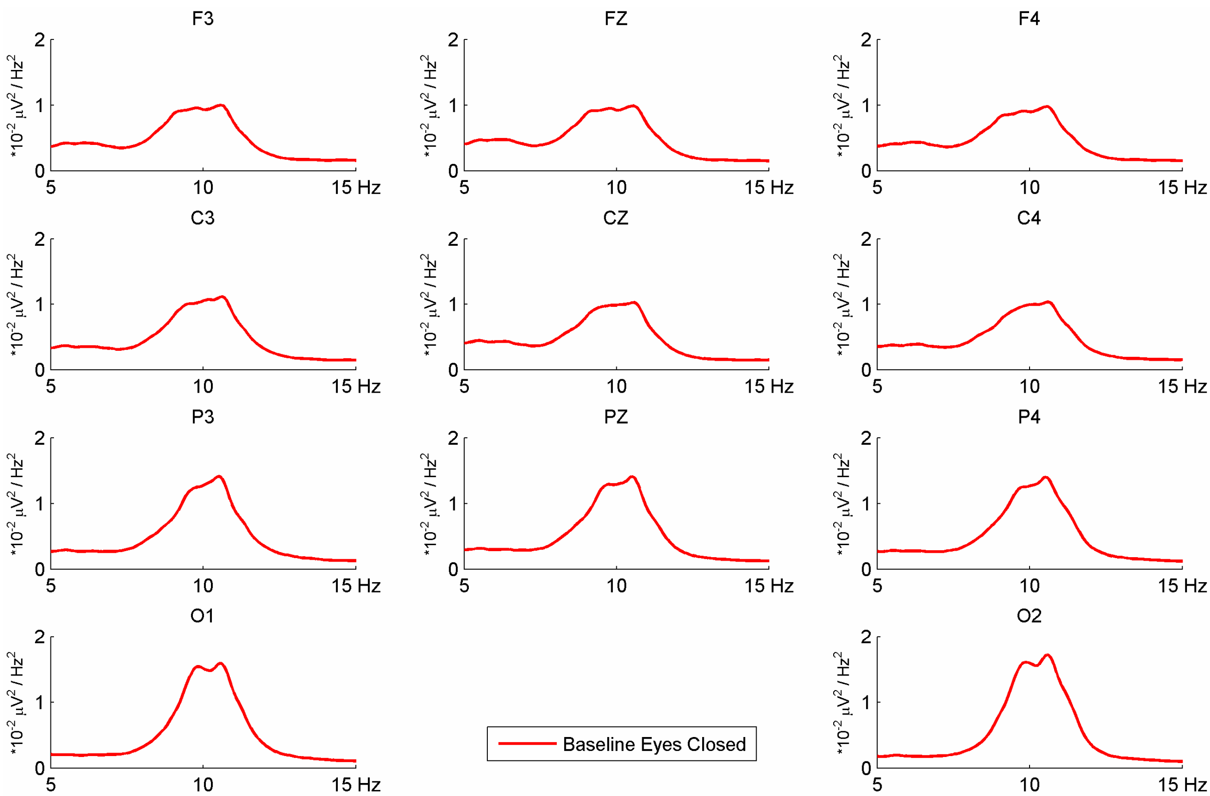

Figure 61. Grand average relative power spectra (n = 20) for the baseline condition with eyes closed in Experiment 1.

Figure 62 shows the grand-average power spectra in the condition in which subjects performed the cognitively demanding task without parallel driving. An increase of alpha power during the more demanding combined story listening task with word detection as opposed to the simple tone detection can be seen. The effect is stronger pronounced over parietal and occipital electrodes and it is equally strong over both electrode hemispheres.

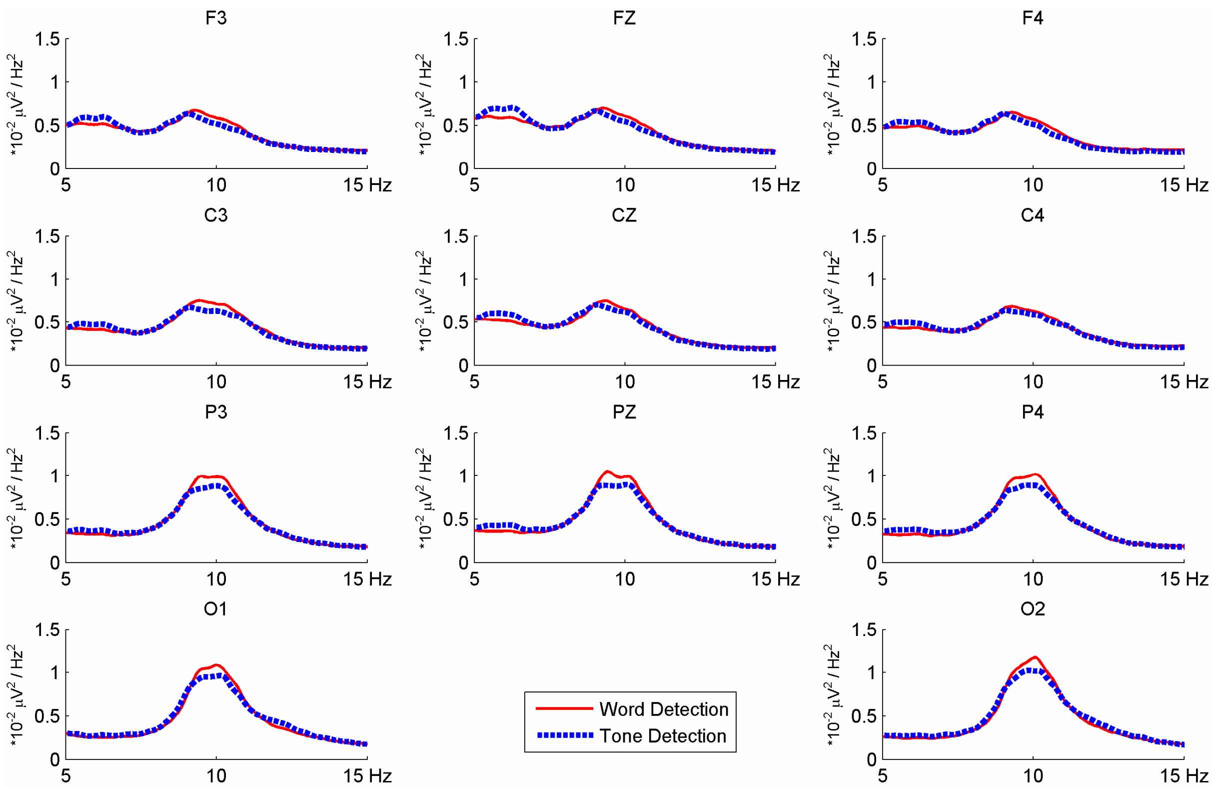

Figure 62. Grand average relative power spectra (n = 20) for the combined story listening task with word detection compared to the tone detection in Experiment 1.

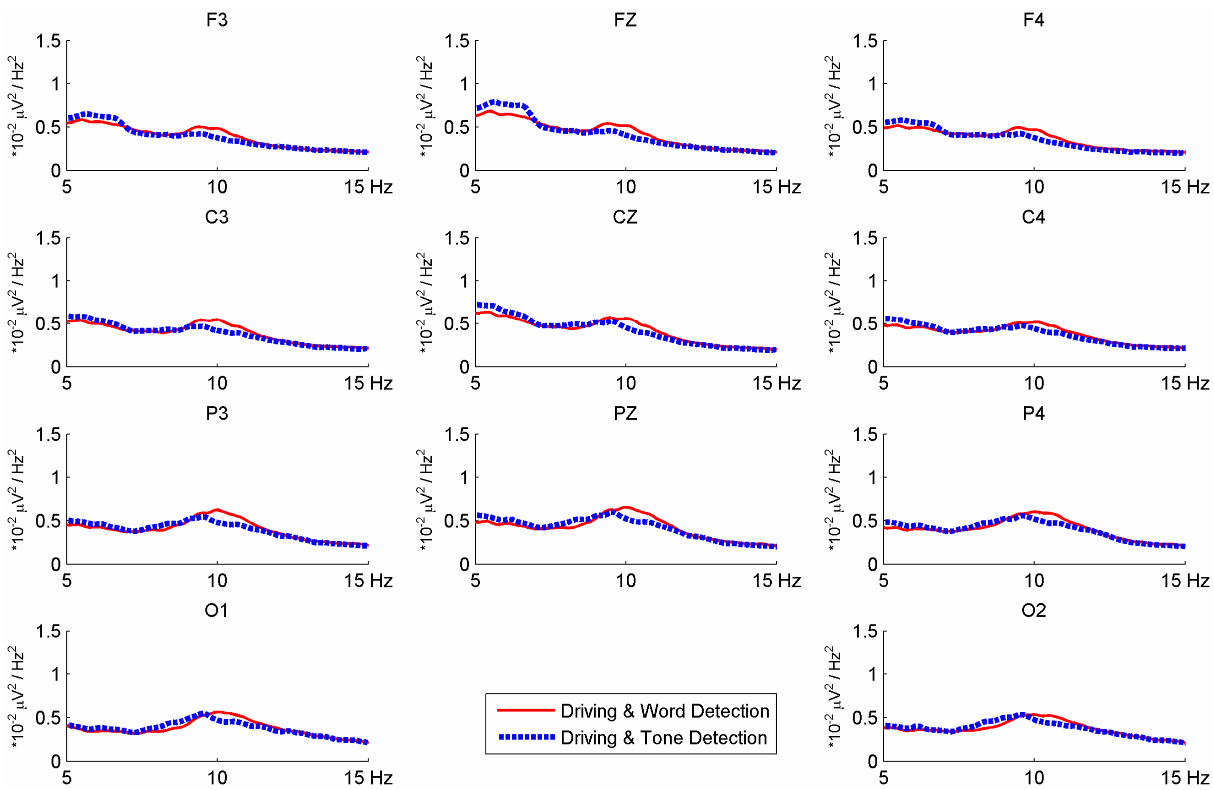

Figure 63. Grand average relative power spectra (n = 20) for the combined story listening task with word detection compared to the tone detection in Experiment 1 while subjects simultaneously performed the Lane Change Test.

A similar pattern of results can be observed in *Figure 63* when subjects performed the two tasks under driving conditions, although the general power level is overall decreased. In addition, the alpha effect is more broadly distributed over all electrodes here. These observations were confirmed by a repeated-measures ANOVA with the three main factors "number of tasks" (word detection vs. LCT and word detection), "task difficulty" (difficult word detection vs. easy tone detection) and "electrode location" (frontal vs. central vs. parietal vs. occipital). As the analysis showed all main effects were significant. The demanding word detection task conditions elicited higher power amplitude values than the simple tone detection tasks ($F(1, 19) = 10.877$, $p < .01$, $\eta^2 = .364$). When comparing alpha in the two experimental conditions in which subjects performed the LCT with the conditions when they performed no simultaneous task an attenuation of alpha can be observed ($F(1, 19) = 51.767$, $p < .001$, $\eta^2 = .732$). In the conditions without LCT alpha power was increasing from frontal to occipital electrode sites ($F(1.58, 30.025) = 24.029$, $pGG < .01$, $\eta^2 = .558$). The interaction between the task difficulty factor and the electrode location factor was not significant ($F < 1$) which indicates that the effect can be observed over all electrode sites. Moreover, the effect is stable regardless if subjects performed an additional driving simulation or whether the detection tasks were performed alone as indicated by the nonsignificant interaction between the task difficulty and the number of task factor ($F(1, 19) = 1.495$). Alpha power is stronger over parietal and occipital than over frontal and central electrodes. However, this difference between electrode locations is more pronounced in the conditions without additional driving task which is supported by a significant interaction between the number of task and electrode location factor ($F(1.395, 26.503) = 24.385$, $pGG < .01$, $\eta^2 = .562$). The effect for number of tasks, namely the reduction in alpha power when performing an additional visuo-perceptive task is highest over occipital electrode sites. The interaction between all three factors was not significant ($F(1.871, 35.545) = 1.2$)

Appendix C: Routes of Experiment 2 and 3

Experiment 2 took place on a two-lane German highway (about 32 km on highway B10 between Esslingen am Neckar and Göppingen). The route is shown in *Figure 64*.

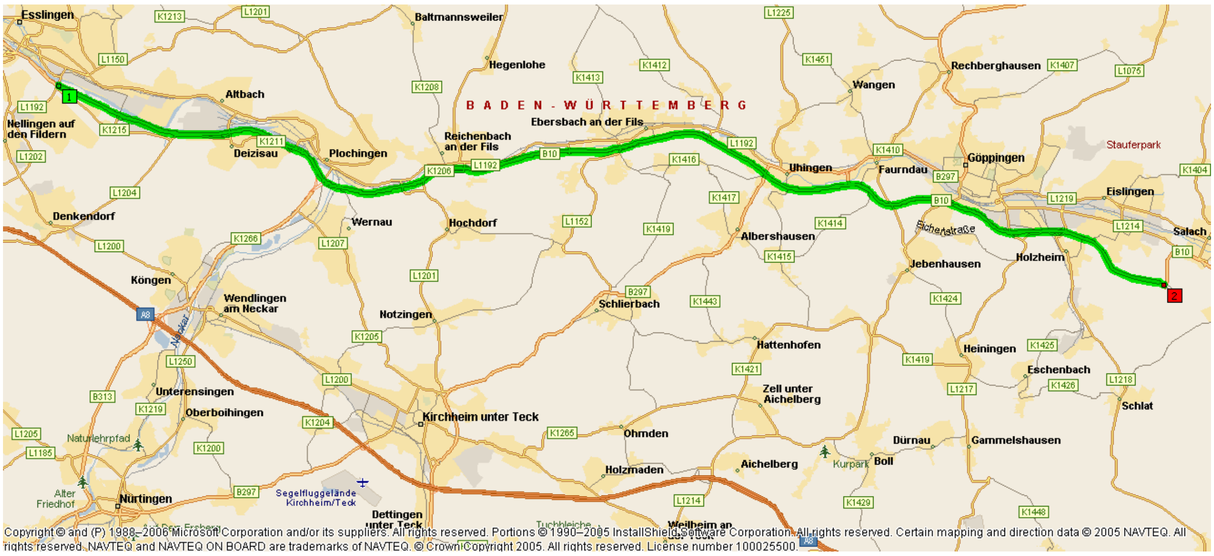

Figure 64. Experimental route of Experiment 2. Subjects started in Esslingen (1) and drove two times towards Göppingen (2) and back.

In experiment 3, subjects drove on dirt roads following predefined routes (see *Figure 65-Figure 72*). Simultaneously recorded GPS data was used to display a sector of about 2.1 km * 1.3 km of a map of the route which was centered on the current position of the car and showed the route. The map was presented on the middle display of the dashboard. For more details please refer to section 5.2.2.2.

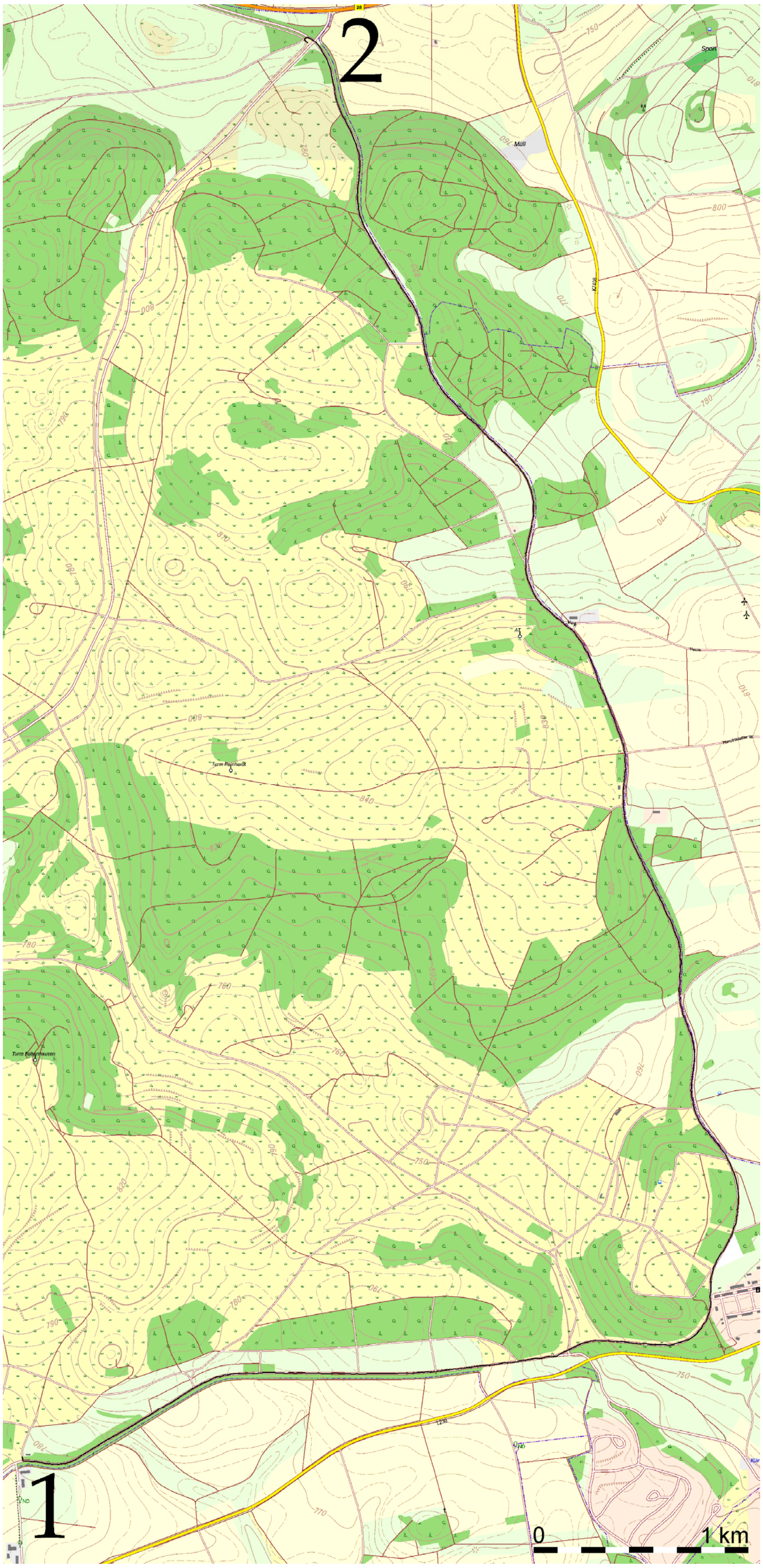

Figure 65. Route of the baseline drive on a broad tarred road in Experiment 3. The baseline drive at the beginning started at Point 1 and ended after ca. 9.6 km at Point 2. For the baseline drive at the end, subjects usually drove a shorter part (ca. 5.9 km) of the route in the other direction.

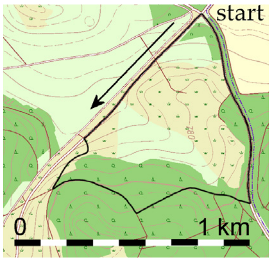

Figure 66. Practice round on dirt roads (ca. 2.9 km) of Experiment 3.

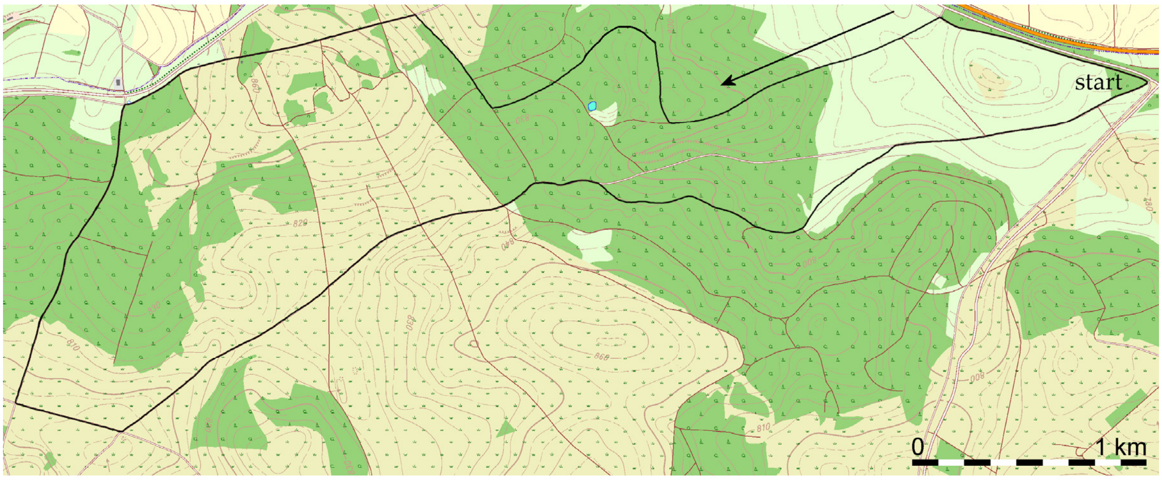

Figure 67. First morning route on dirt roads in Experiment 3 (ca. 11.7 km).

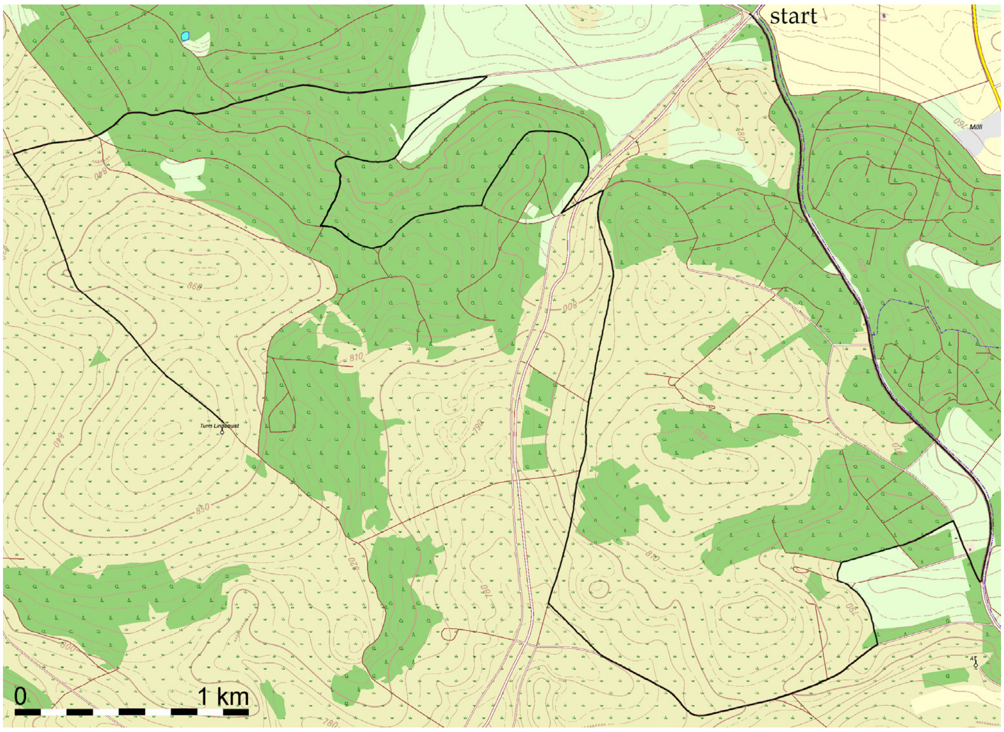

Figure 68. Second morning route on dirt roads in Experiment 3 (ca. 13.7 km).

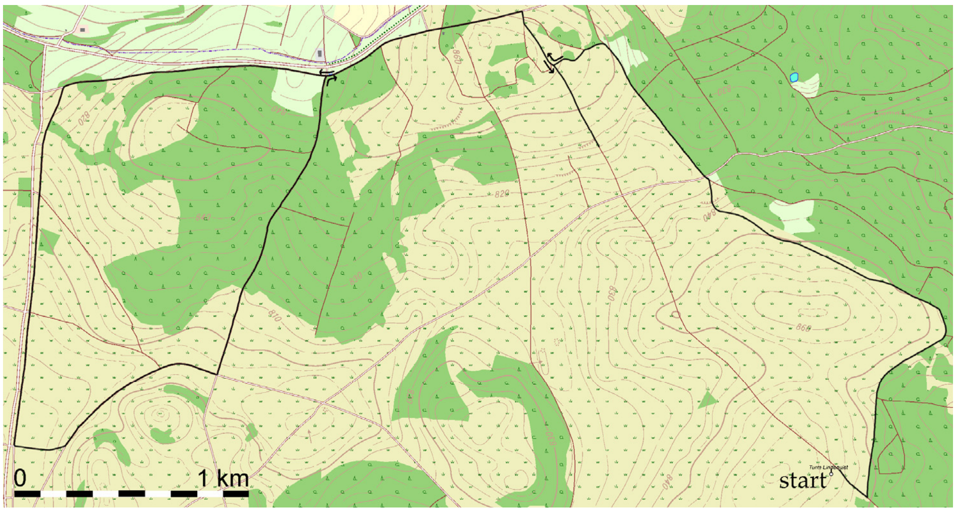

Figure 69. First route of afternoon Set A on dirt roads in Experiment 3 (ca. 10.9 km).

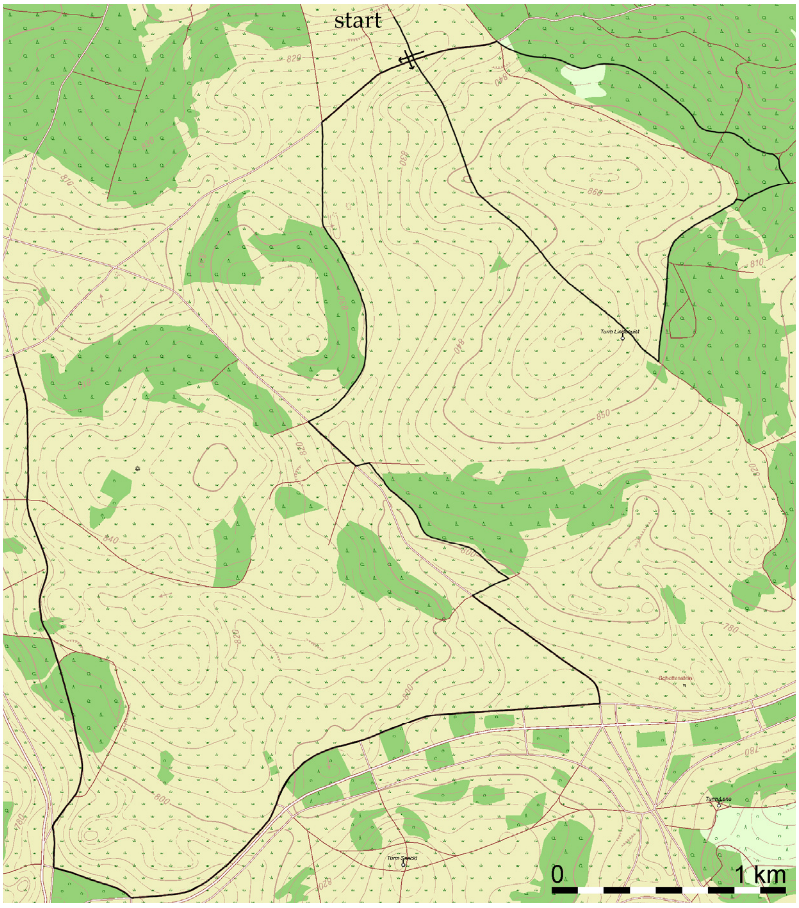

Figure 70. Second route of afternoon Set A on dirt roads in Experiment 3 (ca. 9.4 km).

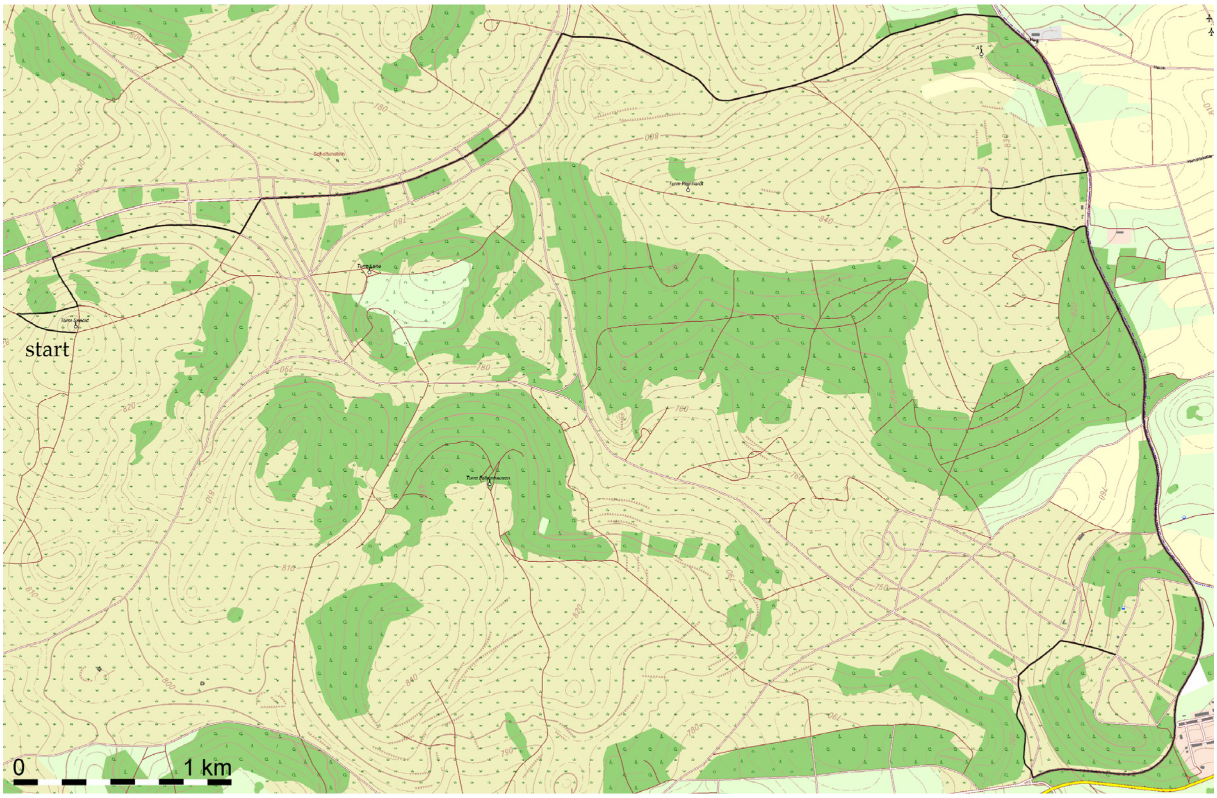

Figure 71. First route of afternoon Set B on dirt roads in Experiment 3 (ca. 11.4 km).

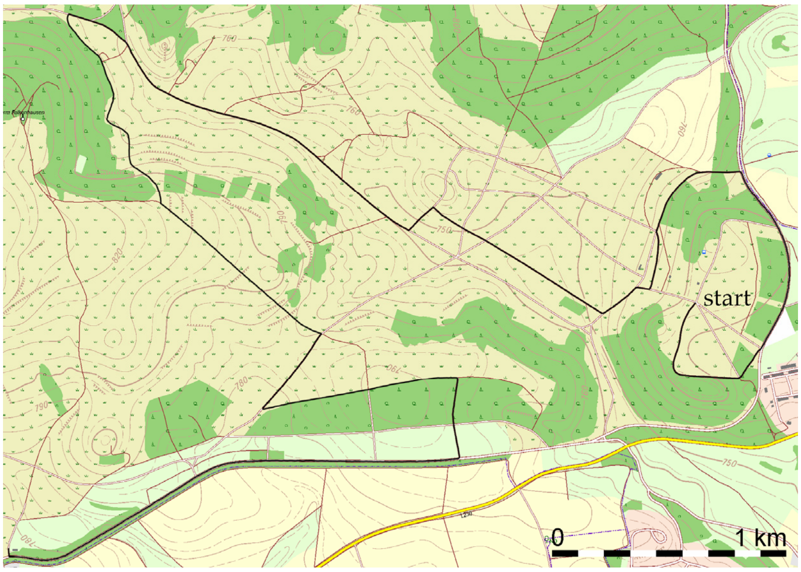

Figure 72. First route of afternoon Set B on dirt roads in Experiment 3 (ca. 10.5 km).