# Monitoring driver’s mental workload for user adaptive aid

Dijksterhuis, Chris

# Monitoring Driver's Mental Workload
# for
# User Adaptive Aid

Cover: André Kikkert, Chris Dijksterhuis
Printed by: Off Page, Amsterdam

ISBN: 978-94-6182-453-0
ISBN: 978-90-367-7109-2
ISBN-Electronic: 978-90-367-7108-5

rijksuniversiteit
groningen

# Monitoring driver's mental workload for user adaptive aid

**Proefschrift**

ter verkrijging van de graad van doctor aan de
Rijksuniversiteit Groningen
op gezag van de
rector magnificus prof. dr. E. Sterken
en volgens besluit van het College voor Promoties.

De openbare verdediging zal plaatsvinden op

donderdag 5 juni 2014 om 14.30 uur

door

**Christiaan Dijksterhuis**

geboren op 27 juli 1980
te Groningen

**Promotor**
Prof. dr. K.A. Brookhuis

**Copromotores**
Dr. ir. L.J.M. Mulder
Dr. D. de Waard

**Beoordelingscommissie**
Prof. dr. M.H. Martens
Prof. dr. M.M. Lorist
Prof. dr. S. Fairclough

# TABLE OF CONTENTS

# Chapter

# 1

# Introduction

Every day, most of us interact with machines to satisfy our needs. For instance, we use phones for communication purposes or to surf on the internet for information. We buy sodas from vending machines, we operate machines at work to satisfy our bosses so that we get paid, and we drive our cars to get to places where we want to be or simply, because we enjoy driving. In other words, we often use machines to get what we need or want, and together, we and the machines perform the necessary tasks.

In this duo task performance situation, one could ask how the machine should behave. At first, this seems a peculiar question to which a straightforward answer can be given. A machine should simply respond to the human who manipulates its interfaces, such as buttons and pedals. In this view of human-machine interaction, a machine is essentially no more than a simple hand tool. However, since the rise of computer technology, machines have become smarter and more autonomous, and therefore, this view is getting increasingly out-of-date. Autopilots can fly airplanes from take-off to landing, manufacturing plants often run largely automatically, driver support systems may prevent unsafe driving behaviour, and while typing this manuscript, subroutines within the word processing software automatically correct spelling errors. In short, computer programmes nowadays assist us with performing mental activities in addition to physical activities, even taking over large parts of it.

But why stop there? In human-human interaction, we continuously monitor each other. Not only to check how a person is performing, but we also keep a close watch on how this person is feeling. Body language indicating stress, overload, fatigue, tiredness, etc. are important cues for telling how someone is doing. It can also indicate how vulnerable this person is to performance breakdowns. As a team member, you would hopefully intervene at some point, for example by assisting that person. In this way we support each other. In the near future, technology will be available that can help us in similar ways. Sensor technology has progressed to the stage that large amounts of information on the internal state of a human being can be acquired, even while this person is engaged with a task. The main remaining challenge is to design the software algorithms capable of interpreting this information in a way that resembles, even to a limited extent, the effectiveness of the human brain. One of the central ideas and assumptions in this thesis is that we will be able to create user adaptive systems with the capacity to monitor our mental state and behave in a way that reflects human-human interaction.

In its most basic form, a user adaptive system can be conceptualised as a continuous feedback loop between the human and the computer part in a human-machine interactive system (Figure 1.1), in which both monitor and react to each other. Of course, this thesis is not the first document in which this concept is presented. An important framework that was directly or indirectly involved throughout this entire thesis is the operator status model (e.g., Hoogeboom and Mulder, 2004) which describes how both behavioural and bodily signals can be acquired by a computer while the user is performing a task in order to determine the mental state of the user, which then leads to specific support actions. Similarly, researchers in the field of adaptive automation have aimed at the same general goal and developed theoretical frameworks during the last three decades, which will be further detailed in Chapter 2.

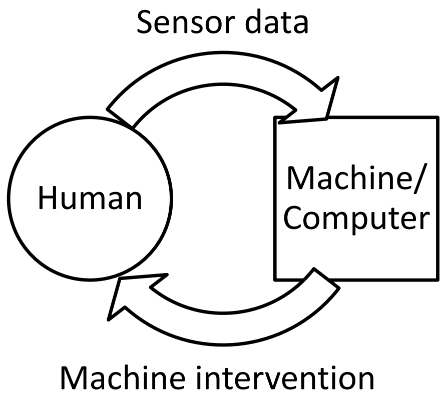



**Figure 1.1**. A schematic view of a user adaptive system. The human user reacts to the task environment, both behaviourally (e.g., task performance) and physiologically (e.g., blood pressure). These reactions are sensed and analysed by a computer. If the computer infers a suboptimal user state, such as mental overload, it decides to intervene, for example by offloading the human user. This intervention is aimed optimizing the user state and so forth.

## 1.1 The mental state, task environment, and support triangle.

As human beings we are able to perceive and respond to a wide variety of mental states in a wide variety of situations. Although this high level of sophistication is the ultimate developmental goal of user adaptive software, the research presented in this thesis is not aimed at meeting such a grand challenge, simply because research in this area has not progressed to the stage that automated mental state assessments and subsequent support actions can be programmed independently from the situational context. Therefore, for the research presented in this thesis specific choices were made with respect to which mental state to focus on, the task environment, and what type of support may benefit the user. Making specific choices however, does not imply that the lessons learned from these studies cannot be extrapolated. In a nutshell, the research presented in this thesis was aimed at investigating how mental workload may be assessed while driving on roads resembling rural conditions (i.e., focussing on the task of keeping a passenger car safely on the road) and how these assessments may be used to avoid extreme workload levels. In this way, this thesis attempts to contribute to the larger vision of advanced human-machine interaction.

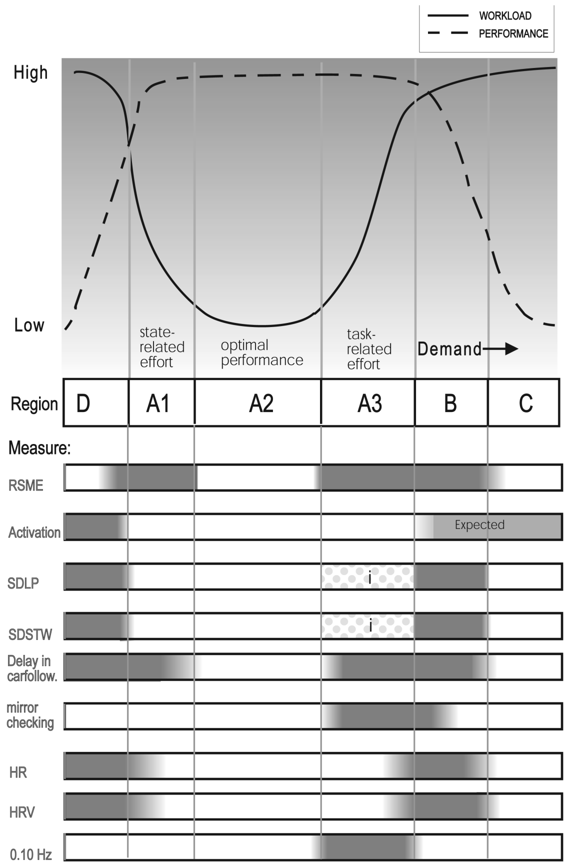



**Figure 1.2.** Workload and performance in six regions (from de Waard, 1996). In region D (D for deactivation), the operator's state is affected and the sensitivity of various measures to detect these regions for the driving task (depicted by the shadings in the horizontal bars). In region A2, performance is optimal: the operator can easily cope with the task requirements and reach a (self-set) adequate level of performance. In the regions A1 and A3 performance remains unaffected but the operator has to exert effort to preserve an undisturbed performance level. In region B this is no longer possible and performance declines while in region C performance is as a minimum level and the operator is overloaded. RSME and Activation are self-report measures. The standard deviation of the car's lateral position (SDLP) and the standard deviation of steering wheel movements (SDSTW) are primary task measures. Lastly, Heart Rate (HR), Heart Rate Variability (HRV), and the 0.10 component of HRV are cardiac measures.

## 1.2 Mental workload

Throughout all studies presented in this thesis, mental workload is a central concept, because of its relevance in understanding how human beings perform in task situations. In 1996, de Waard defined mental workload as 'the reaction to demand' (p. 97) and 'the proportion of capacity that is allocated for task performance' (p. 51). In this thesis of de Waard (1996), the relationship between workload and task demands is visualised as being a U-shaped curve (Figure 1.2), indicating that both low and high demand may lead to high workload. In high demand situations, a person may invest effort to keep up with task demands for a while before performance will degrade (region A3; task-related effort). A low demand situation may invoke a person's natural tendency to rest and decrease alertness. This indicates a decreased capacity to perform a task and trying to stay alert requires a lot of effort (region A1; state-related effort). When task demands are intermediate, mental workload is in fact lowest. In addition, de Waard (1996) visualised the sensitivity of various measures of workload in relation to the driving task (Figure 1.2). The sensitivity that a measure has in reflecting the changes in workload level is an important concept in this thesis as well, in the sense that a minimum level of sensitivity is required to monitor workload changes.

The underlying assumption of mental workload theory is that human beings have one or more limited pools of mental resources that can be directed towards a task (Knowles, 1963; Kahneman, 1973; Wickens, 1984, 2008; Mulder 1986). Mental resources have been 'conceptualised as the availability of one or more pools of general-purpose processing units, capable of performing elementary operations accross a range of tasks, and drawing upon common energy sources' (Hockey, 1997, p. 75). The idea of mental resource scarcity is an important feature in mental workload theory, because it implies that a person may not have enough capacity to adequately perform a task and that some individuals have more capacity than others. For example, mental capacity or maximum effort expenditure may differ from person to person as some people are better trained or more talented. This implies for example, that identical task demands may be more demanding for untrained individuals.

Another important part of workload theory is that resource capacity may change as a function of factors related to our energy household, such as fatigue, since using mental resources requires energy, which is also in scarce supply. For example, a taxi driver near the end of a night shift might not have the energy left to be as alert as during a day shift, indicating a reduced workload capacity. Finally, as human beings we are able to postpone the detrimental effects of depleting our resources and maintain the level of task performance for a while if we feel motivated to do so. However, straining ourselves to protect primary performance may have affective costs such as increases in anxiety, but also compensatory performance costs, such as neglecting secondary tasks (Hockey, 1997,2003).

## 1.3 Maintaining safe lateral control over a vehicle

Although the number of fatalities and severe injuries on roads decreased substantially over the past decades, traffic accidents are still one of the leading causes of death around the world (World Health Organisation, 2009). A major contributor to these accidents statistics are categorised as single vehicle accidents or head-on collisions, especially outside build-up areas (UNECE, 2007), indicating inadequate lane keeping behaviour. In short, this shows that traffic safety may benefit from an adaptive driver support system aimed at improving lane-keeping behaviour, for example, by monitoring signs of suboptimal levels of workload.

The idea of using personalised, adaptive driver support is not new. Already in the 80's, the Generic Intelligent Driver Support research project was aimed at determining the requirements for a 'class of intelligent co-driver systems that are maximally consistent with the information requirements and performance capabilities of the human driver' (GIDS; Michon, 1993, p. 3). The intellectual legacy of these types of pioneering research projects can be found in modern day cars. For example, there are vehicles on the market that can warn a driver or even actively steer to correct the trajectory path if an unintentional lane departure is detected. Another example is Volvo's Intelligent Driver Information System that suppresses an incoming call if the driver is heavily engaged in traffic (e.g., Broström et al., 2006). Developments in the area of advanced driver assistance systems are ongoing and could benefit from knowledge of how to assess the human mental state while being engaged in the driving task.

## 1.4 Thesis outline

Adaptive automation literature has had a large impact on the research in this thesis, and therefore, a theoretical overview of this field is provided in Chapter 2 before moving on to empirical research.

The experiment featured in Chapter 3 has set the driving stage for the other studies. Participants drove through rural road conditions, during which lane-keeping demand was manipulated by changing lane width and oncoming traffic density. The purpose of the experiment was to examine if cardiovascular measures, vehicle parameters, and subjective ratings reflect workload changes. The results of this study confirmed the idea that providing assistance to the driver in this situation may help the driver with maintaining safe control over the vehicle.

In Chapter 4 it is described how drivers experienced and were affected by a support system that was triggered by indications of swerving behaviour within the driving lane, indicating suboptimal workload state. When this happened, the system provided the driver with information with respect to the vehicles position on the road through a head-up display, which the participants could use to improve performance.

In search of alternative mechanisme through which an adaptive support system may affect driving behaviour, the effects of listening to self-rated positive and negative music on lane keeping behaviour, speed choice, and physiological measures during relative low and high demanding rides were investigated and presented in Chapter 5.

In Chapter 6, it is reported how various levels of visuomotor workload, manipulated through speed changes and lane keeping performance targets, were classified through the use of a data-driven machine learning algorithm commonly used in brain-computer interface (BCI) research. This approach allowed determining subject-specific classification models from EEG signals.

For the study that features in Chapter 7, the approach described in Chapter 6 was used to train an EEG-based classification model from calibration data and then apply this model to classify new, incoming EEG data as high, comfortable, or low visuomotor workload. These mental state inferences, together with vehicle indications of worsened lane keeping performance, triggered a cruise control to adapt driving speed, thereby compensating for indications of suboptimal user state.

Finally, in Chapter 8, a general discussion is provided in an attempt to bring together the most important observations and to formulate what can be learned from them in terms of future research directions.

# Chapter

# 2

# Adaptive Automation

## 2.1 Background

One of the oldest documented adaptive automation systems to be found in literature provided aid to a cashier whenever simulated customer queuing times were high (Rouse, 1976). In contrast to assigning a fixed set of tasks to a computer, Rouse proposed a dynamic or adaptive approach to task division (Rouse, 1976; Chu & Rouse 1979). He advocated that, given the fact that there are many tasks that can be successfully performed by both a human and a computer, task performance responsibility should be allocated to whichever partner that has time available. Later, Rouse (1988) describes how this concept had emerged in 1974 out of investigating how to use artificial intelligence in cockpit automation.

During that era there was also a growing concern about the changing role of human workers, operating in ever more technologically advanced environments. Work was shifting from performing tasks manually to monitoring and supervising a process running automatically (Sheridan, 1976a, 1976b; Sheridan & Verplank, 1978). Even though it was clear that automation had a lot of advantages over manual task performance such as: increased capacity and productivity, reduction of small errors, reduction of manual workload and fatigue, relief from routine operations, and more precise handling of routine operations, it also produced a number of unwanted side effects related to the operator being out of the loop. These included, low transparency of what the system does, increased cognitive workload, lowered vigilance, increased boredom, complacency, decreased situation awareness, manual skill erosion, and lower job satisfaction (Wiener & Curry, 1980; Bainbridge, 1983; Wiener, 1989; Endsley, 1995; Billings,1997; de Waard et al., 1999). These drawbacks challenged system's designers to further improve automated systems, for example by keeping the human operator in the loop as much as possible: adaptive automation.

One could ask if human involvement is necessary at all. In 1977, Rouse (p. 391) already argued that 'If the computer can perform every task as well or better than the human, then the computer should do everything'. However, the unlikelihood of this scenario, for example when a computer encounters unexpected situations or because of technology breakdowns, convinced the scientific community that human involvement remained crucial in automation. Interestingly, the conditions formulated by Rouse are more realistic nowadays, as demonstrated for example by self-driving cars (Thrun, 2010). However, assuming that human involvement is still required intermittently, the challenge is how to design a technological system that capitalises on the humans' creative problem solving capacities which computers are still lacking, while at the same time automating simple, routine, tasks as much as possible.

The solution that adaptive automation proposes is adaptive task allocation, which means that tasks are given to and withdrawn from the human operator in real-time, making task allocations dynamic or situation dependent in nature. Initially, the motivation for adaptive allocation of responsibilities was not strictly linked to improving performance as it was stated that it would provide the human operator with the opportunity to step in and perform particular tasks whenever deemed appropriate (Rouse, 1976). However, over time it became accepted to aim allocation policy at optimizing system performance, by keeping human workload or task demand within an acceptable range (Chu & Rouse, 1979; Hancock, 1988; Rouse, 1988; Parasuraman et al., 1992). In addition, since managing task allocation is a task that needs to performed in addition to the primary task, it was argued that the computer should have the authority to at least share the responsibility of initiating task allocation (Rouse, 1977; Scerbo, 1996, 2001; Kaber & Prinzel III, 2006). A system in which the human operator is solely responsible for task allocation is often referred to as adaptable automation (Scerbo, 2001). In general, adaptive automation can be defined as the dynamic allocation of tasks based on the state of the human-task-environment system (Kaber & Riley, 1999).

## 2.2 Theories

Although several theoretical overviews of adaptive automation are given elsewhere, a summary will also be provided here given the importance of the adaptive automation literature for designing the experiments in this thesis (see; Scerbo, 1996; Inagaki, 2003; Sheridan & Parasuraman, 2005; Kaber & Prinzel, 2006; Feigh et al., 2012). To start with, adaptive automation theories may be divided into theories about what the technological system can do and theories about what conditions should trigger the system to adapt. In short, what to adapt and when to adapt.

### 2.2.1 What can be adapted?

Evidently, what can be adapted depends first and foremost on the task at hand. Driving a car is not the same as monitoring screens in a control room, therefore what technology can do to support the user is task dependent. However, in the human factors and ergonomics literature, a number of general theoretical frameworks are present that may help to guide a system designer in thinking about what can be adapted, such as level of automation, dividing tasks into subtasks, and mapping subtasks onto the various stages of the human information processing system.

### 2.2.1.1 Level of automation

An important notion in adaptive automation literature is level of automation. This implies that automation of a particular task does not have to be an all-or-nothing concept and that intermediate levels of automation are also possible. To start with, in 1978, Sheridan and Verplank introduced a list of 'levels of automation in man-computer decision making' (p. 8-17; see Table 2.1). Although this list is often referenced in literature, Sheridan (2000) commented that this list was presented for illustrative purposes mainly and indicated that alternative lists of levels of automation could be constructed (see for example, Endsley & Kaber, 1999).

**Table 2.1.** Two lists of ten levels of automation. On the left hand side, Sheridan & Verplank (1978), on the right hand side, adapted from Endsley & Kaber (1999). LOA = level of automation.

| | Sheridan & Verplank (1978) | | Endsley & Kaber (1999) |
|---|---|---|---|
| 1 | human does the whole job up the point of turning it over to the computer to implement | 1 | Manual control |
| 2 | computer helps by determining the options | 2 | Action support (computer assists performing action, e.g., teleoperator) |
| 3 | computer helps determine options and suggests on, which human need not follow | 3 | Batch processing (computer performs tasks automatically as selected by the operator) |
| 4 | computer selects action and human may or may not do it | 4 | Shared control (computer and operator generate options, the operator selects, performance shared) |
| 5 | computer selects action and implements it if human approves | 5 | Decision support (computer and operator generate options, the operator selects, computer performs) |
| 6 | computer selects action, informs human in plenty of time to stop it | 6 | Blended decision making (computer generates options, computer selects option, operator consents, computer performs) |
| 7 | computer does whole job and necessarily tells human what it did | 7 | Rigid System (computer generates limited number of options, operator consents, computer performs) |
| 8 | computer does whole job and tells human what it did only if human explicitly asks | 8 | Automated decision making (computer and operator generate options, computer selects and performs) |
| 9 | computer does whole job and tells human what it did and it, the computer, decides he should be told | 9 | Supervisory control (computer generates, selects and performs. Operator monitors and might intervene by shifting the system to a lower LOA). |
| 10 | computer does whole job if it decides it should be done, and if so tells human, if it decides he should be told | 10 | Full automation (the human is completely out of the control loop) |

Expanding on this idea, the various levels of automation can also be described as a one-dimensional continuum of automation degrees (Parasuraman et al., 2000), where the term assistance could be used when most of the task is performed by the human operator, including support types that are not readily included in the concept of automation, such as providing information, warnings, and advice (Flemisch et al., 2008). Conversely, at the other end of the scale the term automated could be used when most of the task is automated

#### 2.2.1.2 The human information processing system

Another way of categorizing automation is to map its tasks onto the human information processing system, which was proposed by several authors (see Table 2.2).

**Table 2.2.** Stages of the human information processing system

| | **Sheridan (2000)** | **Parasuraman et al. (2000)** | **Endsley & Kaber (1999)** |
|---|---|---|---|
| 1 | Acquire information | Sensory Processing | Monitoring (system status) |
| 2 | Analyse & Display | Perception/Working memory | Generating (options for action) |
| 3 | Decide Action | Decision making | Selecting (deciding on a particular option) |
| 4 | Implement action | Response Selection | Implementing (carrying out the chosen option) |

Within this framework, acquisition automation involves supporting the human sensory processes, such as highlighting important information that is present on a computer screen. Analysis automation involves the support of cognitive functions. For example, a computer may help to integrate sensory data by combining several information sources into one measure. Decision automation means helping a human operator with selecting from action alternatives, for example by indicating the most efficient course of action. Finally, action automation refers to the execution of a selected action (Parasuraman et al., 2000).

Furthermore, Parasuraman et al. (2000) suggested that each stage of the information processing system may be automated to a certain degree. For example, system A can have a moderately high level of acquisition automation, but low levels of automation for the other stages, while system B has a high level of automation across all stages (See Figure 2.1).

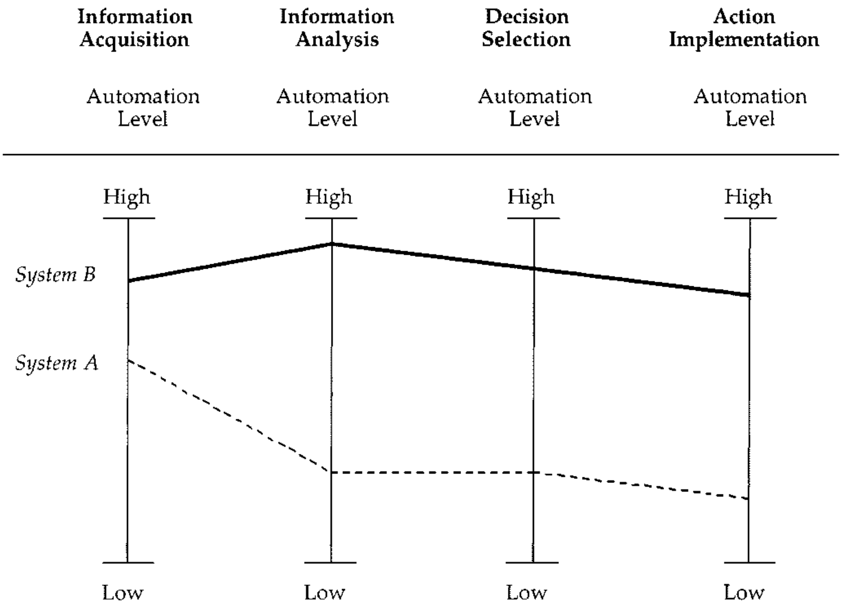



**Figure 2.1.** Levels of automation for independent functions of information acquisition, information analysis, decision selection, and action implementation. Examples of systems with different levels of automation across functional dimensions are also shown (From Parasuraman et al., 2000, © 2000 IEEE).

### 2.2.1.3 Beyond task allocation

Most literature on adaptive automation focuses on dynamic task allocation as a way to load or offload a human operator. However, the adaptive possibilities of a system do not have to be limited to (sub)task allocations. Already in the early eighties, transforming a task, such as changing the level of abstraction on a display was mentioned in addition to dynamically allocating complete tasks and task partitioning (Rouse & Rouse, 1983; see also Feigh et al., 2012). Another example of how a system could transform the current level of automation is changing intensities of warnings when the user seems distracted instead of just switching on and off a warning system. Changing the adaptation timing parameters as a result of human state inference would be another option. For example, a lane keeping driver assistance system might always be activated but could be made to react sooner and stronger when the computer notices signs of fatigue.

The term *meta adaptation* could be used to indicate changes in the way that a system adapts over time as it gains experience with how a user responds. The term meta in this context is borrowed from Miller & Funk (2000) who argued that an adaptive system should be able to communicate what it infers and accept feedback from the human pilots which it uses to improve itself. This kind of communication is commonplace in human-human interaction and they referred to this as meta-communication. In driver assistance systems research, this type of high-level adaptation has also been mentioned. For example, in GIDS it was argued that a navigation device could gradually provide less and less route information as the driver gains experience with driving a particular route (progressive modification; Michon, 1993). Interestingly, in software engineering, it is also recognized that interactions between the user and the system may be differentiated in terms of timescales for system adaptation. The slowest adaptive time scale can be referred to as evolutionary computing (e.g., Serbedzija & Fairclough, 2009). Meta adaptations will not only optimise the way the computer interacts with its user, but may be a necessary step in the development of such systems since human responses may change over time. The last reason why meta adaptations could be crucial for real-world feasibility are behavioural adaptations by the human user. In transport psychology, behavioural adaptations (such as driving less carefully when using a seat belt) following the implementation of safety measures are known phenomenon and usually mean that road users feel safe enough to change their behaviour in a way that leaves the safety potential partly unfulfilled (e.g., Brookhuis & de Waard, 2004). As phrased by Hancock & Verwey (1997): 'The system developer is faced with designing for humans who themselves often respond in unexpected ways because they themselves are adaptive' (p. 499).

### 2.2.2 When should a system adapt?

A computer-initiated adaptive system should infer when it is time to change level of automation. In general, the sources that the system may use for this purpose are the task environment (indicating task demands), human behaviour (such as task performance), and the internal state of the human operator (e.g., Kaber & Riley, 1999; Hoogeboom & Mulder, 2004). In adaptive automation literature, specific methods for invoking automation have also been formulated (see Table 2.3).

**Table 2.3** Automation invocation methods

| Parasuraman et al. (1992) | Inagaki (2003) |
|---|---|
| Critical-Event Logic | Critical-Event Strategies |
| Dynamic assessment of Pilot Mental workload<br>Dynamic Psychophysiological assessment | Measurement-based Strategies |
| Pilot Performance Models | Model based strategies |

#### 2.2.2.1 Critical-event driven

Critical-event logic means that aid is only triggered after the occurrence of a specific external events, such as an emergency situation. For example, the European Commission adopted two proposals to ensure that by 2015, all new cars will automatically call the emergency number in the event of a serious accident (European Commission, 2013). Although an adaptive system based on this strategy is directly linked to events in the real world, it is insensitive to actual pilot performance or workload. In other words, it will provide support regardless of whether or not the system's user needs or desires it when the event occurs (Parasuraman et al., 1992; Inagaki, 2003; Sheridan & Parasuraman, 2005).

#### 2.2.2.2 Performance driven

A performance driven adaptive system strategy may help to tailor the system to the individual user. It will react by changing the level of automation if that particular user performs below threshold levels. However, since an important goal of an adaptive system is to prevent performance degradation, a reactive system may not be sufficient. Also, it ignores that maintaining adequate performance levels may at times have negative impacts on the user's wellbeing. As is clear from workload related paradigms such as the operator functional framework (Hockey, 1997, 2003), human beings may exhaust themselves by straining effort expenditure to protect primary task performance in demanding situations. Performance protection is important for dealing with short bursts of task demand. However, when exposed to longer periods of high workload, this will require increasingly more effort (task-related effort, see Figure 1.2), which may have affective costs such as increases in anxiety, but also compensatory performance costs, such as neglecting secondary tasks. Hockey refers to this situation as a compromised system state, in which the individual is more vulnerable to a performance breakdown (Hockey, 1997, 2003). Frequent exposure to high demanding situations may elicit even more serious health issues in the long term. For instance, it has been suggested that repetitive activation of a cardiovascular defence reflex, which leads to an immediate increase of the heart rate and blood pressure, may also lead to hypertension in the long run (e.g., Johnson and Anderson, 1990).

#### 2.2.2.3 Physiologically driven

Using physiological data has the potential to create a more proactive system. Since straining effort expenditure as described above has a neurophysiological base, the ability to reliably classify workload from neurophysiological data could be used to offload a person, before performance effects become apparent. Monitoring galvanic skin response, heart rate, blood pressure, and electroencephalography are examples of signals that have been used to assess effort investment (Pope et al, 1995; de Waard, 1996; Prinzel III et al., 2003; Wilson & Russell, 2007; Mulder et al., 2009; Dijksterhuis et al., 2013).

However, the fundamental problem of these measures is the complex relationship between mental states, such as workload, and their associated biometrical variables (Fairclough, 2009). To start with, a unique one-to-one relationship between the biometrical variable and the psychological construct would be ideal for an adaptive system, but is very rare. A many-to-one relationship is more complicated as several signals are needed to infer a mental state. In a one-to-many relationship, one biometrical signal is sensitive to more mental states. Lastly, a many-to-many relationship means that many signals are in fact sensitive to many mental states, making it very hard to reliably classify a mental state (Cacioppo et al., 2000). In general, in adaptive automation we have to do with this many-to-many relationship. In addition, there is a problem of generalising the relationship between a biometrical measure and a user state outside the laboratory as a mapping may not hold true or is different in the real world were conditions are less controlled.

#### 2.2.2.4 Operator modelling

Most adaptive automation research effort has probably been directed towards developing operator modelling types of systems, resulting in a series of associate research programs (see Rouse et al., 1990 for example). Most notably, starting in the mid-eighties, was the Pilot Associate in which adaptive automation was applied to air-to-air combat missions (e.g., Banks and Lizza, 1991). Documented spin-off programmes include the Rotocraft Pilot Associate, the Crewman's Associate for battle tanks, and an associate for Unmanned Aerial Vehicles (Pechacek, 1991; Miller & Hannen, 1999). An important feature of these associates was that they tracked the pilot's actions and then matched these onto an existing knowledge base containing numerous task situations and expected human actions. Automation was activated if there was sufficient mismatch between the current activities and the expected activities.

#### 2.2.2.5 Other concerns with respect to timing

External events such as high task demands and internal events such as stress and high workload may indicate a need for support and subsequently determine the system's behaviour. However, when implementing a user adaptive system, several other factors need to be taken into account (Scerbo, 1996; Hoogeboom & Mulder, 2004; Mulder et al., 2008 ).

A system's responsiveness for example, refers to how fast a system responds. This depends partly on the time period over which data needs to be collected before a reliable mental state inference can be made. On the one hand, using longer time windows may increase reliability. On the other hand, a slow responding system may not be effective simply because any extra assistance is no longer required, indicating a potential speed-accuracy trade off. Yet again, rapidly changing levels of automation might not be effective as an operator may have difficulties adapting to a different level of automation, necessitating dead band periods during which additional support actions are not allowed. Also, maintaining automation awareness in a rapidly changing system may become a burden to the human operator. Another concern is to avoid large oscillations in the system behaviour by limiting the loop-gain, for example by preventing large changes of the task as a result of task allocations. These considerations imply that a direct link between a human state and level of support provided by the system may not always be possible or sensible.

# Chapter

# 3

# Steering Demand and Mental Workload

Chapter 3 is based on Dijksterhuis, C., Brookhuis, K.A., & de Waard, D. (2011). Effects of steering demand on lane keeping behaviour, self-reports, and physiology. A simulator study. Accident Analysis and Prevention, 43, 1074-1081. DOI: 10.1016/j.aap.2010.12.014

### Abstract

In this study a driving simulator was used to determine changes in mental effort in response to manipulations of steering demand. Changes in mental effort were assessed by using subjective effort ratings, physiology, and the standard deviation of the lateral position. Steering demand was increased by exposure to narrow lane widths and high density oncoming traffic while speed was fixed in all conditions to prevent a compensatory reaction. Results indicated that both steering demand factors influence mental effort expenditure and using multiple measures contributes to effort assessment. Application of these outcomes for adaptive automation is envisaged.


## 3.1 Introduction

The task of keeping the vehicle in the driving lane is a relatively easy but continuous and important part of driving. Though easy, the continuous character increases the likelihood of errors with time on task. Therefore, it is hardly surprising that accidents preceded by inadequate steering behaviour make up a large proportion of accident statistics. In 2003 for example, one third of approximately 855,000 accidents involving personal injury or death in Canada, France, Germany, and the Netherlands were categorised as single vehicle accidents (24%) or head-on collisions (9%). Moreover, when outside build-up areas the percentages of these categories increase to 36% and 13% respectively, summing the number of lateral control related accidents involving personal injury or death outside build-up areas up to about 172,000 (UNECE, 2007).

The dangers of insufficient lateral control have also been illustrated in literature; often combined with speeding. In an investigation into the nature of heavy vehicle fatal crashes in Victoria, Australia (n=61), it was revealed that over a third of these accidents involved leaving the roadway on a straight road and most crashes (44) involved just one vehicle (Brodie et al., 2009). In a study by Jamson et al. (2008), expert drivers observed driving scenarios and judged occasional lane departures to reduce safety. In a UK sample of 1185 vehicle occupant fatalities, 44% involved going loss of control in a bend or curve (Clarke et al., 2010). Finally, in a survey amongst road users involved in accidents in Germany, 14% of 284 accidents on urban roads and 53% of 190 accidents on non-urban roads were categorised as lane departure crashes (Staubach, 2009). The large proportion of accidents related to lateral control indicates the relevance of research into factors associated with inadequate steering behaviour.

The vast majority of traffic accidents can be traced back to human error and workload related problems. That is, when mental effort mobilisation in service of the driving task is not enough to ensure adequate driving performance (e.g., Brookhuis & de Waard, 2010). Assessment of mental effort investment can therefore be a valuable tool to investigate potential dangerous or demanding driving situations. Moreover, automated mental effort assessment in real-time could be used by advanced driving assistance systems (ADAS) to trigger driving assistance (e.g., Mulder et al., 2009). Currently, commercially available lane keeping assistance systems in passenger cars use road markings to function and nearly all are triggered in case those markings are crossed without using the indicator signal. However, indications of uncomfortable levels of mental strain, even if performance thresholds are not yet violated may warrant the triggering of a warning or a corrective steering action by an ADAS as a preventive measure.

At least two broad categories of automatically detecting changes in mental effort expenditure can be distinguished; performance assessment and assessing psychophysiology. Ideally, more effort investment would always lead to improved performance, in which case performance assessment would suffice to assess effort expenditure. However, a linear relationship between these two variables does not always exist (e.g., de Waard, 1996). For instance, someone may already operate at a maximum performance levels and extra effort expenditure will not result in performance increase. Similarly, changes in performance levels do not necessarily reflect changes in mental effort. Performance may deteriorate as a result of increasing task demands, without any change in effort expenditure. For example, if a driver chooses to conserve energy by not increasing workload level to near maximum capacity and accept a lower performance standard (e.g., Hockey, 1997, 2003). This suggests that performance assessment by itself is not adequate for mental workload assessment.

In addition to measuring performance levels, drivers' mental workload may be assessed automatically by measuring physiological reactions. In this way, changes in effort expenditure may be revealed even if task behaviour does not change. Measuring cardiovascular activities during periods of mental effort investment has been a subject of investigations for several decades. Especially heart rate variability (HRV) has been shown to react to performing a mentally demanding task (Mulder, 1992). Moreover, HRV (specifically centred around 0.1 Hz) was concluded to be a sensitive index of 'invested mental effort' (Boucsein & Backs, 2000; de Waard et al., 2008; de Waard et al., 2009; Mulder, 1986). In addition to cardiovascular indications, several other physiological signals may be used to infer mental workload (for an overview of signals that can be measured in a driving simulator see for example Brookhuis & de Waard, 2010).

The main purpose of the present study is to investigate the sensitivity of and relations between performance, subjective, and physiological indices of mental effort expenditure for several levels of steering demand, mainly for the purpose of future ADAS development. For this, a simulator study was conducted in which driving behaviour on several narrow lane widths were compared with the normal lane width (3 metres) on a standard Dutch two-lane rural road while confronted with low density oncoming traffic. To further increase steering demand, (the perceived) manoeuvring space of the driver was decreased by nesting a period of high density oncoming traffic in each lane width section. As a behavioural adaptation to increasing steering demands, drivers would be expected to decrease speed (e.g., de Waard et al., 2004). However, since assessing effort investment under high load is an objective in the presents study, it was decided to preset and fix the driving speed of the simulator car to prevent such a compensatory speed reaction.

Although lane width is a basic factor in all driving research, literature studies that have used this an independent variable, either physical or optical, are scarce (e.g., Godley et al., 2004; Lewis-Evans & Charlton, 2006; Rosey et al., 2009). Also, making clear comparisons of effects of lane width across studies is complicated given the wide variety of possible confounding factors such as shoulder width, total road width, road type, road curvature, speed limitations, and the presence of pavements, trees, or buildings in the direct vicinity of the road (e.g., Van Driel et al., 2004). As a result it is not entirely clear what behavioural results to expect from the current investigation.

Similarly, a stream of traffic travelling in opposite direction is often included in driving simulator research as standard part of the simulated road environment, although literature on a comparative study on different intensities of traffic flow on the opposite lane was not found by the authors. Meeting a vehicle in the opposite lane usually results in a lateral displacement towards the road edge (Räsänen, 2005; Rosey et al., 2009). Hypothetically, high density oncoming traffic may therefore result in a further shift toward the road edge.

The lane width and oncoming traffic conditions in this experiment are designed to increase task demand on the lane keeping task from normal to high. It can therefore be expected that drivers will increase mental effort expenditure, as indicated by changes in the standard deviation of the lateral position, increased subjective effort ratings, and changes in physiology associated with maintaining driving performance standards. However, there is also a possibility that under conditions of high steering demand, drivers are no longer able or willing to further increase lane keeping performance. In this case, a more diffuse pattern of steering behaviour, ratings, and physiology is expected.

## 3.2 Method

### 3.2.1 Participants

A total of 22 males and 8 females were recruited through posters placed around the University of Groningen and were paid 20 Euros for participating. Age ranged from 22 to 39 years (mean = 26.6; SD = 4.1) and participants had held their licence for 4 to 20 years (mean = 7.9; SD = 3.4). Self-reported total mileage ranged from 15,000 to 500,000 km (median = 45,000; interquartile range (IQR) = 92,500) and the reported annual mileage for the past two years ranged from 4,000 to 40,000 km (median = 8,250; IQR = 14,750). None of the participants reported using prescribed drugs that might affect driving behaviour.

### 3.2.2 Design

Four levels of lane width (3.00, 2.75, 2.50, and 2.25 m) were created by dividing a driving circuit into four main sections of uninterrupted road, stretching out for 9.1 km on average per section (about 7 minutes of driving time), winding through rural scenery, and separated by small towns (see Figure 3.1 for a screenshot of the driving environment). Oncoming traffic was generated with a random interval gap between 7 and 13 seconds, resulting in 10 passing private vehicles (width: 1.75 m) per minute on average. However, on the last 1.4 km of each section, the interval gap shortened to one to two seconds, resulting in an oncoming traffic density of 40 cars per minute. Moreover, to increase lateral demand even further, half of the passing vehicles were small lorries (width: 2.26 m). In this way, a short high density oncoming traffic section was nested into each main section. Each participant drove the simulated car (width: 1.60 m) through all sections creating eight periods of interest for each participant: 4 (lane width) × 2 (oncoming traffic density). When driving on one of the main sections, speed was controlled by the simulator and set to 80 km/h to prevent the potential compensatory reaction of slowing down by the participants.

Road curvature was designed to be similar within and between sections. Each section consisted mainly of faint curves (85% curves, 15% straight) with a radius of 382.0 m for all curves. The average curve length was 193.3 m (SD = 23.5) which corresponds to a mean angle of 29.0 degrees (SD = 3.5). The road surface was marked on the edges by a continuous line (20 cm), in the centre by a discontinuous line (15 cm), and outside the edges a soft shoulder (width: 2 m) was present for the entire experimental route. Except for oncoming traffic, no objects were present on or in the direct vicinity of the road.

**Figure 3.1.** Impression of the roadway environment as seen by the driver on the front screen of the simulator.

### 3.2.3 Simulator

The (fixed-base) driving simulator used consists of a mock-up car with functional pedals, clutch, steering wheel, safety belt, indicator and handbrake. The simulator runs on ST Software© which is capable of simulating fully interactive traffic. The three computers dedicated to the simulator compute the road environment and traffic at 30Hz+, which are displayed on three 32-inch plasma screens and provide a total view of the driving environment of 210°. A detailed description of the driving simulator software used can be found in Van Winsum & Van Wolffelaar (1993).

### 3.2.4 Procedure

Participants were informed about the experiment in general terms upon arrival and signed an informed consent. After this, the participants filled out a short demographic questionnaire before ECG electrodes and a respiration belt were placed on the body. Participants were then given the opportunity to get used to the simulator by driving around on a two-lane practise road (lane width: 3.00 m) without any other traffic. After this, ECG and respiration were recorded for three minutes to provide physiological baseline values while the participant was requested to relax. Thereupon, it took about 36 minutes for each participant to complete the driving circuit. This time included the time it took the participants to drive through the build-up areas (lane width: 3.00 m) in between the experimental main sections. In each town, participants received the instruction to pull-over and park the car. This break lasted two minutes during which the participants were asked to mark the effort and risk rating scales for both the low and high oncoming traffic conditions outside the build-up area. In this way, four ratings were requested from the participants in each of the four towns. Finally, another three minute physiological measurement baseline period was recorded when the simulation was shut down, directly after rating the last low and high oncoming traffic sections.

All participants drove the circuit in clockwise direction. Therefore, given a certain start location, the sequence in which participants were exposed to the lane width conditions was fixed (e.g., 3.00 m, 2.50 m, 2.25 m, and 2.75 m when participants started in the town indicated in the top of Figure 3.2). However, start location was counterbalanced over towns to minimise sequence effects.

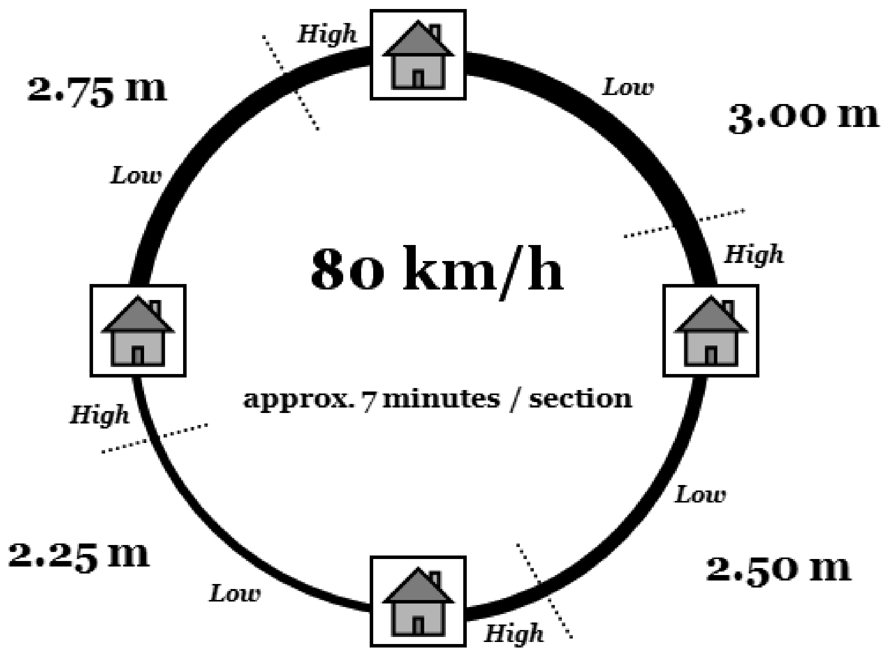



**Figure 3.2.** Abstract representation of the route driven. Houses symbolise towns, increasing line thickness represents increasing lane width, and high/low represents high/low oncoming traffic density.

### 3.2.5 Measures

During the two-minute break in between experimental main sections, a rating on the one-dimensional Rating Scale Mental Effort (RSME; Zijlstra, 1993) was requested, separately for the first segment: low density oncoming traffic, and the last segment: high density oncoming traffic. The RSME ranges from 0 to 150 and participants may use these digits to rate experienced effort. In addition, several effort indications are visible alongside the scale which may further guide the participant in marking the scale. Indications start with 'absolutely no effort' (RSME score of 2) and end with 'extreme effort' (RSME score of 112). Experienced risk was rated on an identical scale, except that the word "effort" was substituted by "risk".

Lateral Position (LP) was sampled at 10 Hz and is defined as the difference in metres between the centre of the participant's car and the middle of the (right hand) driving lane. Positive LP values correspond to deviations toward the left hand shoulder and negative values correspond to deviations toward the right hand shoulder. The sampled LP values were used to calculate mean LP, the standard deviation of LP (SDLP), and the proportion of the time that any part of the vehicle was outside the lane edges for each of the eight experimental sections.

### 3.2.6 Physiology

Two physiological signals were sampled, both at 250 Hz. Firstly, the electrocardiogram (ECG) was registered using three Ag-AgCl electrodes, which were placed on the sternum (the common electrode) and on the right and left side between the two lower ribs. R-peaks in the ECG signal were detected online with an accuracy of 4 ms and these were used to create inter-beat interval (IBI) time series. Artifacts in the IBI time series were corrected automatically using the CARSPAN spectral analysis program, visually inspected for deviations and thereupon processed for spectral analysis (Mulder, 1992). In this way, mean heart rate (HR), and the power of heart rate variability (HRV) in the mid frequencies (0.07 – 0.14 Hz) were derived. Variations in the mid-frequency band of 0.07 – 0.14 Hz reflect variations in mental effort. Secondly, a respiration signal was recorded by means of a respiration belt (Twente Medical systems, RespitraceTM principle). Spectral plots of the respiration signal where then used to visually determine the main respiration period in each condition.

3.2.7 Analysis

The data were analysed using the General Linear Model for Repeated Measures analysis in SPSS (version 17.0). For every dependent variable, the overal main and interaction effects were analysed by using the multivariate approach for repeated measures. For a detailed assessment of lane width effects, trend analyses were carried out by using polynomial contrasts. Furthermore, pairwise comparisons for both main and interaction effects were carried out by using SPSS simple contrasts; resulting p-values were multiplied in accordance to the Bonferroni method.

## 3.3 Results

### 3.3.1 Subjective ratings

To begin with, risk scores were regressed on RSME scores and the results indicate a positive relationship ($\beta = 0.83$, $t(238) = 7.34$, $p < .001$). Also, risk explains a significant proportion of variance in RSME scores ($R^2 = 0.695$, $F(1, 238) = 545.43$, $p < 0.001$). Since risk and RSME scores are highly correlated and yield the same repeated measures test results, the focus in this section will be on reporting average scores of the RSME.

Mean mental effort scores range from just over the 'almost no effort' mark on the RSME (17 for 3.00 m width / low density traffic) to just over the 'some effort' mark (46 for 2.25 m width / high density traffic; see Figure 3.3). Within this range, several significant effects were revealed.

**Table 3.1.** Multivariate test results for subjective effort and experienced risk. Width: lane width; Traffic; oncoming traffic density.

| **Subjective Ratings** | | | | | | | |
|---|---|---|---|---|---|---|---|
| | | **RSME** | | | **Risk** | | |
| *Effect* | *df 1, 2* | **F** | **p** | **$\eta^2$** | **F** | **p** | **$\eta^2$** |
| **Traffic** | 1, 29 | 59.88 | <0.001 | 0.674 | 62.63 | <0.001 | 0.684 |
| **Width** | 3, 27 | 6.38 | 0.002 | 0.415 | 12.75 | <0.001 | 0.586 |
| **Traffic x Width** | 3, 27 | 7.48 | 0.001 | 0.454 | 6.38 | 0.002 | 0.413 |

Overall, high density oncoming traffic conditions were rated as more effortful than the low traffic conditions. Also, lane width has a main effect on effort ratings which increases as lane width decreases (linear trend: $F(1,29) = 18.6$, $p < 0.001$, $\eta^2 = 0.39$). However, the differences in effort ratings between the various lane width levels are not equal for both traffic density conditions, creating an interaction. As can be seen in Figure 3.3, the effort increase from the 3.00 m to the 2.25 m lane width is small for the low traffic condition, while a larger increase can be seen during the high traffic conditions (from 26 to 46 RSME points).

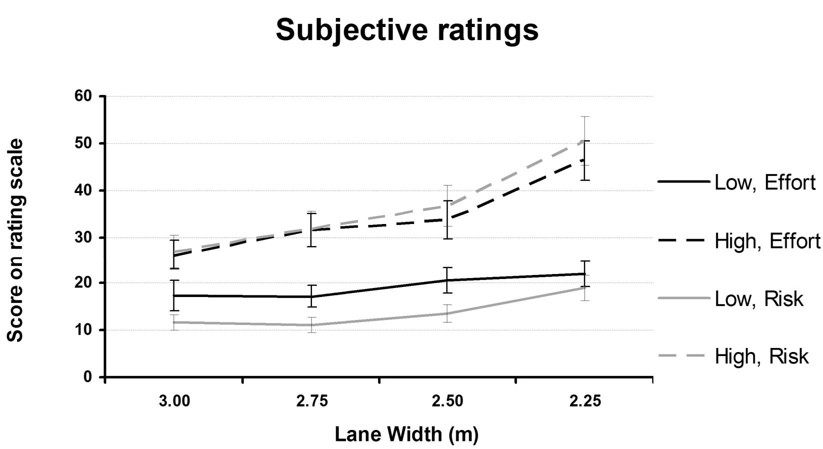



**Figure 3.3.** Reported Mental Effort (RSME) and perceived Risk. Low/High represent low/high oncoming traffic density. Error bars represent the standard error. Maximum score for both mental effort and experienced risk is 150.

### 3.3.2 Vehicle parameters

Test result of the vehicle parameters are shown in Table 3.2.

**Table 3.2.** Multivariate test results for Lateral Position, the Standard Deviation of the Lateral Position, and Driving time over the lane's lines. Width: lane width; Traffic; oncoming traffic density.

| Vehicle Parameters | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | LP | | | SDLP | | | Proportion over lines | | |
| *Effect* | *df 1, 2* | F | p | $\eta^2$ | F | p | $\eta^2$ | F | p | $\eta^2$ |
| **Traffic** | 1, 29 | 382.19 | <0.001 | 0.929 | 23.13 | <0.001 | 0.444 | 69.36 | <0.001 | 0.705 |
| **Width** | 3, 27 | 3.88 | 0.020 | 0.301 | 15.12 | <0.001 | 0.627 | 95.14 | <0.001 | 0.914 |
| **Traffic x Width** | 3, 27 | 4.91 | 0.008 | 0.353 | <1 | ns | 0.055 | 43.66 | <0.001 | 0.829 |

#### 3.3.2.1 Lateral Position

Lateral position (LP) represents the distance between the centre of the car and the centre of the driving lane. In all conditions, mean LP is negative (see Figure 3.4) indicating that the participants' preferred a position on the right hand side of the lane centre. On average, participants drove 0.07 m towards the shoulder during periods of low density oncoming traffic. During high oncoming traffic density, this distance increased to 0.31 m. The statistical analysis revealed that mean LP was not the same during all levels of lane width. However, analyses of lane width contrasts by doing pairwise comparisons revealed that none of the LP differences between levels of lane are significant, although two differences are 'marginally significant' (using an alpha of 0.1). These were the LP differences between the 3.00 m and 2.50 m lane width conditions and the 2.75 m and 2.50 m conditions (0.03 m and 0.03 m respectively; $F(1,29) = 7.65$ and 12.48, $p = 0.059$ and 0.094, $\eta^2 = 0.21$ and 0.30).

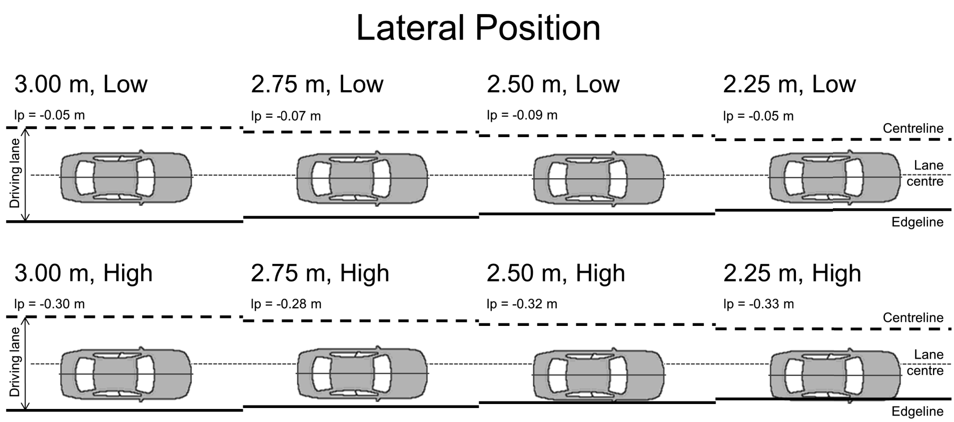



**Figure 3.4.** The mean lateral position of the vehicle in the driving lane. Lateral position is defined as the distance between the centre of the car and the centre of the driving lane. Negative values denote a position right of the lane centre. The centre of the driving lane is indicated by a thin dashed line and the centre of the car by a thin solid line overlaying each car icon. 3.00 m, 2.75 m, 2.50 m, and 2.25 m are driving lane widths, Low/High represent low/high oncoming traffic density.

The interaction between lane width and oncoming traffic (see Table 3.2) indicates that the LP differences found amongst the four levels of lane width is different for the low density oncoming traffic condition when compared to the high density oncoming traffic condition. When taking a closer look at Figure 3.4 and using the standard 3.00 m lane width as a reference, it can be seen that the largest effect difference between traffic densities is in the

2.75 m lane. Compared to the 3.00 m lane width, participants drove 0.02 m more towards the opposite driving lane in the low traffic condition while driving 0.06 m more towards the right shoulder in the high traffic condition (marginally significant: $F(1,29) = 3.3$, $p = 0.08$, $\eta^2 = 0.10$).

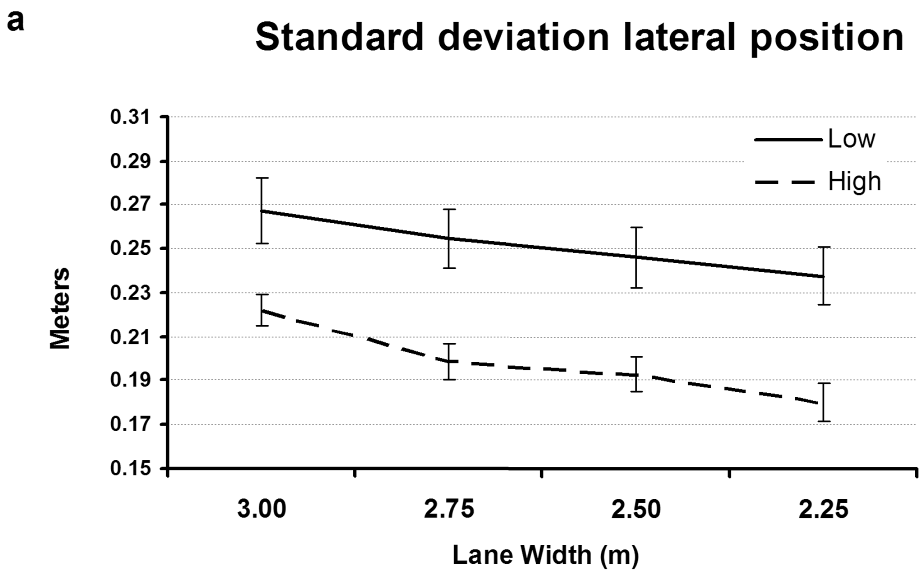
Please note that the LP values presented in Figure 3.4, do not indicate how close the car was to crossing the lane delineation and as such does not inform how close a driver would have been to a dangerous situation in a real car. In Figure 3.4, this is shown by depicting the car (width: 1.60 m) in proportion to each lane width. Naturally, total manoeuvring space in the driving lane decreases as lane width decreases. For example, a mean LP of 0.05 m (3.00 m width/ low density traffic) means that participants had 0.65 m manoeuvring space on their right hand side and 0.75 m on their left hand side in the 3.00 m width condition. The same LP value in the 2.25 m width condition shows that participants had 0.275 m lane space on their right and 0.375 m on their left hand side. Similarly, the low LP value of -0.33 m in the narrowest lane condition during high traffic meant that participants preferred to drive 1 cm on the shoulder and maintained an average distance of 0.66 m between their car and the opposite lane.

3.3.2.2 SDLP

The standard deviation of the lateral position (SDLP) shows clear effects of both lane width and oncoming traffic (see Figure 3.5a). During periods of high density oncoming traffic, mean SDLP was 0.05 m lower than SDLP during the low traffic conditions (0.20 m vs. 0.25 m). Moreover, the difference between high and low traffic remains constant while the overall SDLP decreases about 0.01 m for each 0.25 m reduction in lane width. The consistent decrease of SDLP was further reflected by polynomial contrast analysis, which revealed a linear relation ($F(1,29) = 39.7$, $p < 0.001$, $\eta^2 = 0.58$). Interaction effects were not present in the SDLP data.

3.3.2.3 Driving time over the lines.

In Figure 3.5b, the proportion of the driving time is displayed that participants allowed any part of the vehicle to cross the road surface markings on either side of the driving lane. Overall, the proportion ‘over the lines’ is 0.15 higher in the high traffic conditions. Also, there is an overall increase of driving time outside the lane’s delineation which is reflected by significant linear and quadratic trends (linear: $F(1,29) = 286.3$, $p< 0.001$, $\eta^2 = 0.91$; quadratic: $F(1,29) = 42.4$, $p<0.001$, $\eta^2 = 0.59$).



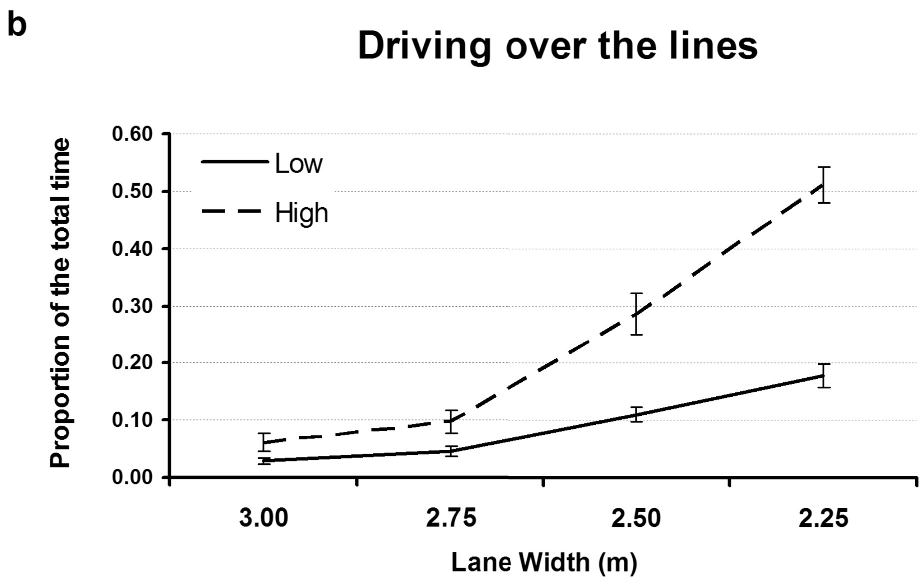




**Figure 3.5.** Standard Deviation of the Lateral Position (a) and the proportion of the total time driven outside the lane's delineation (b). Low/High represent low/high oncoming traffic density. Error bars show the standard error.

Moreover, the lines in Figure 3.5b diverge as lane width decreases and this is exhibited by an overall interaction effect. In the 3.00 m lane width condition, the proportions for both traffic conditions are closest together; 0.01 for the low and 0.05 for the high traffic conditions. In the 2.25 m lane condition the difference between both traffic densities expands to 0.34 as proportions peaked at 0.18 for the low and 0.51 for the high traffic condition.

Last, the proportion driving time over the lines was not equally divided over both sides of the driving lane. In the low oncoming traffic conditions, participants drove twice as much on the shoulder than in the opposite lane. This share varied between 50% in the 3.00 m lane width condition and 72% in the 2.50 m condition. During high density oncoming traffic, driving on the shoulder accounted for 99.5% of the total time over the lines; regardless of lane width.

#### 3.3.2.4 Crashes

Throughout the experiment, six participants caused a total of six accidents as they hit a vehicle in the opposite driving lane. All of these accidents happened on the most narrow driving lane and four of them during high oncoming traffic density. To investigate potential differences between the six crash involved individuals and the 24 other participants, the repeated measures analysis which results were reported earlier were repeated using crash involved individuals as a between subject factor. As it turned out, crash involved individuals rated experienced risk 14 points lower on average over all conditions ($F(1,28) = 6.1$, $p = 0.020$, $\eta^2 = 0.18$). Moreover, the low density conditions were rated 6 points lower while the high density traffic conditions were rated 22 points lower by the crash involved individuals, creating an interaction ($F(1,28) = 5.9$, $p = 0.022$, $\eta^2 = 0.17$). SDLP was also different for the two groups. Overall, SDLP was higher for the crash involved individuals (0.26 m vs. 0.22 m; $F(1,28) = 5.3$, $p = 0.028$, $\eta^2 = 0.16$), resulting in a main effect of the group factor. However, the SDLP difference between groups was more distinct for the low oncoming traffic density (0.31 m vs. 0.24 m) than the high density conditions (0.22 m vs. 0.19 m; marginally significant interactions: $F(1,28) = 3.6$, $p = 0.068$, $\eta^2 = 0.11$). No group effects were revealed for any of the physiological measures.

### 3.3.3 Physiological measures

Test result of the (transformed) physiological measures are shown in Table 3.3.

**Table 3.3.** Multivariate test results for Heart Rate (HR), Heart Rate Variability in the midband frequencies (HRV-M), and Respiration. Width: lane width; Traffic; oncoming traffic density. HRV-M was Ln-transformed before statistical testing. Driving (vs. BL) represents the difference between the average (8) driving vs. the average (2) baseline periods. Significant effects (p< .05) are shown in **bold**.

| **Physiological measures** | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | **HR** | | | **HRV-M** | | | **Respiration** | | |
| *Effect* | *df 1, 2* | F | p | $\eta^2$ | F | p | $\eta^2$ | F | p | $\eta^2$ |
| **Traffic** | **1, 29** | **5.76** | **0.023** | **0.166** | **12.78** | **0.001** | **0.306** | <1 | ns | 0.017 |
| **Width** | 3, 27 | 1.42 | ns | 0.137 | 1.44 | ns | 0.138 | 1.845 | ns | 0.170 |
| **Traffic x Width** | 3, 27 | <1 | ns | 0.044 | <1 | ns | 0.967 | 1.434 | ns | 0.138 |
| **Driving (vs. BL)** | 1, 29 | 3.03 | 0.092 | 0.095 | 7.238 | 0.012 | 0.200 | 34.29 | <0.001 | 0.542 |

#### 3.3.3.1 Effects during driving (experimental effects)

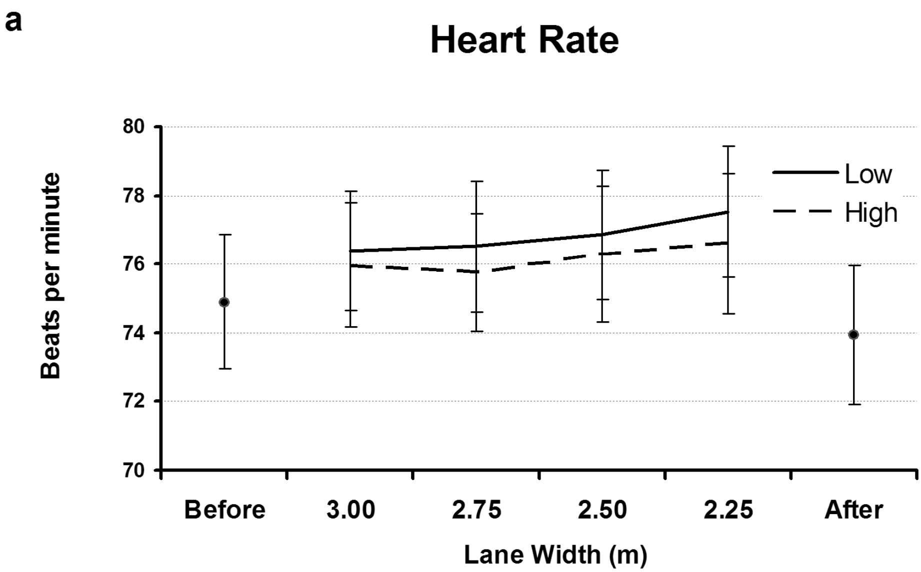
Neither HR, HRV-M, nor respiration period differed between the four lane widths. However, an increase in oncoming traffic density did reveal effects on physiology. For HR, more oncoming traffic resulted in a slight but significant decrease from 76.8 to 76.2 bpm. (see Figure 3.6a). In addition, power in the midband frequencies of the HR (HRV-M) decreased; from 7.04 to 6.72 (mMI2 (ln-transformed)) on average (see Figure 3.6b).

#### 3.3.3.2 Driving vs. baseline

The HR difference between driving and the baseline periods is not significant, although a trend (at alpha = 0.1) was revealed (Table 3.3). That is to say, the average HR during driving was 76.5 bpm, which is 2.1 bpm higher than the average HR during the baseline periods (see Figure 3.6a). Driving-baseline effects are clearer for the other physiological measures. HRV-M decreases when driving (6.88 vs. 7.29 mMI2 (ln-transformed) for the baseline; see Figure 3.6b) and participants breathed faster compared to the average baseline value (3.64 vs. 5.53 seconds main respiration period).

#### 3.3.3.3 Low and high density oncoming traffic sequence

The nested one-minute period of high oncoming density oncoming traffic at the end of each level of lane width creates the possibility of sequence effects. To investigate possible data trends prior to the high traffic conditions, all physiological and vehicle data were cut into one-minute segments. Although there is variation between each minute for all lane widths, no trends that were consistent over all four levels of lane width could be observed for HR, HRV-M, LP, or SDLP.



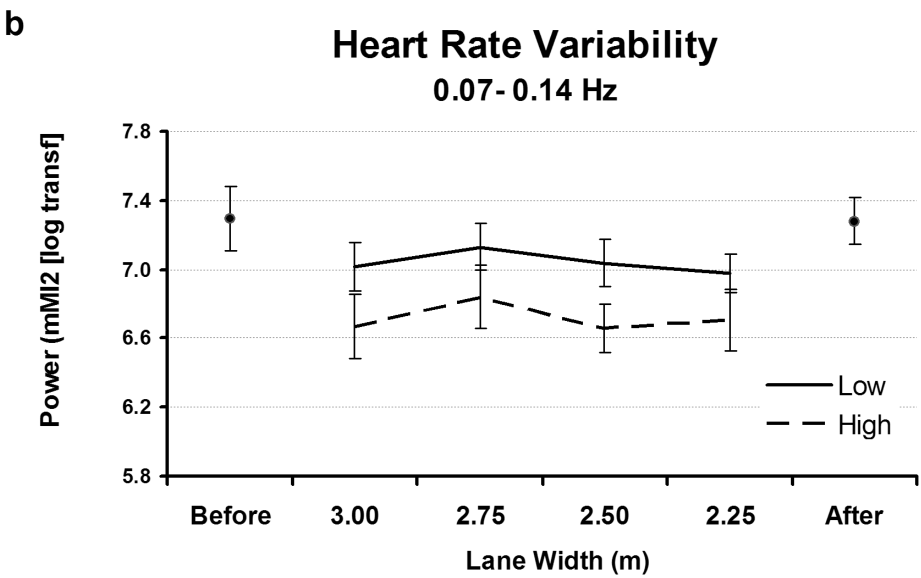




**Figure 3.6.** Heart rate (a) and heart rate variability in the mid frequency band(b; HRV-M). A decrease in HRV-M reflects more effort investment. Low/High represent low/high oncoming traffic density. Before/After represent a three-minute baseline period before/after the ride. Error bars show the standard error.

## 3.4 Discussion and Conclusions

The main aim of the experimental manipulations was to create several levels of driving demand to investigate how steering behaviour is affected by these manipulations and to associate these changes with mental effort as reflected by driving performance, self-ratings, and physiological reactions. Before discussing mental effort or workload issues, steering behaviour and possible signs of performance degradation will be discussed in which Summala's Multiple Model Monitor (Summala, 2005) will be used to interpret performance effects.

### 3.4.1 Steering behaviour

The results indicated significant effects of both lane width and oncoming traffic density on all steering parameters (LP, SDLP, and proportion of the driving time over the lane edges). As expected, the increase in oncoming traffic during the last minute of each road section resulted in a deviation from the lane centre towards the shoulder. Next, in contrast to Lewis-Evans & Charlton (2006) and Rosey et al. (2009), narrower lane widths were not associated with a LP closer to the road centre. Regretfully, a number of factors between these previous studies and the current study differ and therefore hinder formulating a clear explanation. These factors include lane width, shoulder width, background terrain, the implementation of oncoming traffic, and road elevations which all have a potential influence on road position behaviour (e.g., Van Driel, 2004). Furthermore, with regard to Rosey et al. (2009), there is a difference in methodology. In Rosey et al. (2009), LP is measured as the distance between the car and the road centre instead of the distance to the lane centre, which by definition decreases as lane width decreases in case the vehicle maintains a position in the centre of the lane.

When taking a closer look at the position on the road with respect to the lane edges, it is striking that the mean distance between the side of the car and the shoulder continues to decrease during heavy oncoming traffic. Participants even preferred to drive slightly over the lane markings in the narrowest lane condition. In terms of safety margins, this means that the comfort zone of the participants must have been partly on the road shoulder. The large proportion of driving time on the shoulder confirms this. On the one hand, these behavioural observations are surprising as driving on a soft shoulder in real road conditions is dangerous. On the other hand, given the heavy traffic on the opposite lane, attempting to maximise the distance to the opposite lane might be an effective coping strategy. In the driving simulator used, feedback signals that are usually associated with driving on a soft shoulder, like loss of traction and bumpy car movements were not present, although a rumbling sound alerted the participants when wandering off-road for more than 0.40 m. This made the simulated shoulder look like a soft shoulder but part of it felt like a hard shoulder. Participants might have learned to anticipate on this and use the 'extra' manoeuvring space. Given this possibility, the large proportion of driving time that participants crossed the lane markings cannot readily be interpreted as a mark of performance degradation. In contrast, from the perspective of the drivers, it is more likely that performance levels were maintained since they maximised the safety margin to the opposite lane.

If lateral position reflects the centre of a safety zone, then SDLP might show the width of this zone. Participants swerved less as lane width decreased, which is in agreement with previous findings (Godley, 2004; de Waard et al., 1995). Furthermore, SDLP also decreased under conditions of heavy oncoming traffic and this effect adds to the lane width effect. Apparently, drivers did cope with all levels of lateral demand by decreasing the zone in which they drove by improved steering performance. Nevertheless, six accidents occurred during the investigation on the narrowest lane width. The latter raises the issue of validity. Especially when driving behaviour is observed that would have been disastrous in real-world driving it makes sense to doubt the validity of the simulator in these conditions. With respect to this issue, a distinction needs to be made between absolute and relative validity (e.g., Blaauw, 1982; Bella, 2008; Godley et al., 2002; Törnros, 1998). Törnros (1998) suggested that for a driving simulator to be a useful research tool, relative validity is satisfactory. That is, if the researchers seek to investigate effects of independent measures, rather than determining numerical values of driving behaviour. In case of the present study, participants were exposed to a 2.25 m driving lane width for several minutes while speed was set at 80 km/h. Although maintaining lane keeping performance under these conditions may be viewed as challenging for the perceptual-motor system, the occurrence of six accidents out of a total of 256 road sections is an indication that this result is unlikely to reflect real world accident statistics. However, this does not imply that the results of the current experiment are not valid in a relative sense.

The performance of the crash involved participants show the worse steering behaviour than the other participants and the crash involved individuals might therefore be more prone to accidents than the others. As it turned out, the crash involved participants rated experienced risk lower than the others and displayed a higher standard deviation of the lateral position. Speculatively, these individuals did not maintain the same safety margins as the others. In summary, the crash occurrences should be viewed as an indication of bad lateral control which may be the result of both driving simulator characteristics and the steering skills of the crash involved participants.

### 3.4.2 Effort

Decreased SDLP requires more intensive monitoring of the road and frequent steering wheel corrections and can therefore be interpreted as more demanding. Subsequently, it can be concluded that more effort was put into the steering task as lane width decreased and oncoming traffic density increased. These main effects are confirmed by subjective ratings, although an interaction was also revealed. Participants did not indicate an increase in effort or risk perception as a result of increasing lane width during the low density oncoming traffic periods. In de Waard's workload model (de Waard, 1996), workload is minimal and performance is maximal when task demands are intermediate, whereas workload increases when demands either increase or decrease. Given the range of the scales (0-150) it could be argued that subjective effort and experienced risk are almost absent and not sensitive to changes in steering demand during periods of light oncoming traffic which in turn is an indication that drivers were performing optimally and relative effortless. However, during periods of heavy oncoming traffic, both ratings show a clear increase as a result of decreasing lane widths, indicating that steering demands were at intermediate to high levels.

Experienced risk reflected lateral demand in the same way as subjective effort in terms of statistical significance, a finding which is consistent with previous research (e.g., Lewis-Evans & Rothengatter, 2009). However, effect sizes as reflected by partial eta squared is an indication that experienced risk might be more sensitive to changes in steering demand.

Effect sizes on physiology were smaller than effects sizes on subjective ratings, although significant effects of oncoming traffic on heart rate and heart rate variability were found. Heart rate variability in the midband frequencies was lower on sections with high oncoming traffic, indicating more mental effort expenditure (e.g., Boucsein & Backs, 2000). Interestingly, lower heart rate variability was accompanied a slightly lower heart rate during high density oncoming traffic. Although any explanation of this unexpected result is speculative, we suspect that differences between the current study and other studies might be related to the nature of the steering (tracking) task, which could be the topic of future research.

### 3.4.3 Conclusions

For every increasing level of lateral demand, extra effort was mobilised in service of the steering task as indicated by a decrease in SDLP. No clear trend was observed for lateral displacement of the vehicle as a result of lane width variations, although an increase in oncoming traffic was associated with a position to the right of the lane centre. Although lateral control, as measured by SDLP, improved for every level of steering demand, subjective ratings were only sensitive to different levels of lane width under conditions of high demand. A reduction of heart rate variability was associated with high oncoming traffic, indicating that this measure is less sensitive in detecting measuring effort expenditure as a result of lane width level than the other measures.

Based on these findings, several suggestions can be made for developing ADAS systems. ADAS systems are envisioned that dynamically support the driver based on the principles in the field of adaptive automation (e.g., Miller & Parasuraman, 2007; Parasuraman & Riley, 1997; Scerbo, 1996). For instance, detection of high effort expenditure may be utilised by ADAS systems to minimise distractions from other in-vehicle devices which compete for the driver's attention, such as phones or navigation systems. These secondary systems could be automatically delayed or temporarily shut down, if the drivers' physiology or steering performance indicate high or low effort expenditure in service of the steering task. In addition, the primary control task of keeping the vehicle in the driving lane could be supported based on mental effort assessment instead of the lane departures which are the current standard for lane assist systems. Since other performance measures such as SDLP or physiological measures are likely to show signs of increased or decreased effort expenditure before lane departures occur, lane assist systems could be used to prevent more serious driving errors to happen in the first place. In conclusion, the results from the present research may be used as a starting point for the development of ADAS systems which assess mental effort to trigger driving support and determine what type of support is most appropriate.

## Acknowledgement

The research presented in this chapter was supported by the Commission of the European Communion, Project No. 215893 REFLECT, Responsive Flexible Collaborating Ambient.

# Chapter

# 4

# A Performance Based Adaptive Driver Support System

Chapter 4 is based on Dijksterhuis, C, Stuiver, A, Mulder, L.J.M., Brookhuis, K.A., & de Waard, D. (2012). An adaptive driver support system: user experiences and driving performance in a simulator. Human Factors, 54, 772-785. DOI: 10.1177/0018720811430502

**Abstract**

*Objective:* The aim of this study was to test the implementation of an adaptive driver support system. *Background:* Providing support might not always be desirable from a safety perspective, as support may lead to problems related to a human operator being out of the loop. In contrast, adaptive support systems are designed to keep the operator in the loop as much as possible by providing support only when necessary. *Method:* A total of 31 experienced drivers were exposed to three modes of lane-keeping support: nonadaptive, adaptive, and no support. Support involved continuously updated lateral position feedback shown on a head-up display. When adaptive, support was triggered by performance-based indications of effort investment. Narrowing lane width and increasing density of oncoming traffic served to increase steering demand, and speed was fixed in all conditions to prevent any compensatory speed reactions. *Results:* Participants preferred the adaptive support mode mainly as a warning signal and tended to ignore nonadaptive feedback. Furthermore, driving behaviour was improved by adaptive support in that participants drove more centrally, displayed less lateral variation and drove less outside the lane's delineation when support was in the adaptive mode compared with both the no-support mode and the nonadaptive support mode. *Conclusion:* A human operator is likely to use machine-triggered adaptations as an indication that thresholds have been passed, regardless of the support that is initiated. Therefore supporting only the sensory processing stage of the human information processing system with adaptive automation may not be feasible. *Application:* These conclusions are relevant for designing adaptive driver support systems.

## 4.1 Introduction

Enabled by technology, fully automated cars driving around in real traffic are no longer science fiction. In 2010, Google announced that it is testing self-driven cars using trained safety drivers inside the cars to take over control when necessary (Lee, 2010; Thrun, 2010). At the time of the announcement, seven test cars had driven a total of 140,000 miles with occasional human intervention and 1,000 miles without any human involvement (Markoff, 2010). Only one car was involved in an accident, when it was rear-ended while waiting at a traffic light. In the same announcement, it was pointed out that an automated car allows its passengers to spend their time more efficiently, improves safety, and increases road capacity, since self-driving cars respond faster than humans, have a 360° perception, do not get distracted or fatigued, and so on.

Although sitting back and relaxing while the automated system handles vehicle control is appealing, it represents a fundamental change of the driver's role from an active driver to a passive supervisor, potentially resulting in less safety (Brookhuis & de Waard, 2007). This change is similar to what has been seen in other high-tech task environments. Although full automation has many advantages, including reduction of small errors and better handling of routine operations (Wiener & Curry, 1980), there can be serious risks to performance when a manual takeover is required, as a result of reduced situation awareness, manual skill erosion, and late detection of automation failures (Bainbridge, 1983; de Waard, Van der Hulst, Hoedemaeker, & Brookhuis, 1999; Endsley, 1995). The aforementioned example shows that even the impressive driving performance shown by the Google test cars does not allow a driver at present to sit back and fully engage in other activities.

To prevent "out-of-the-loop" problems, it has been proposed to automate (sub)tasks only when support is needed, known as adaptive automation (Byrne & Parasuraman, 1996; Miller & Parasuraman, 2007; Mulder, Dijksterhuis, Stuiver, & de Waard, 2009; Parasuraman & Riley, 1997; Scerbo, 1996, 2001). In contrast to those in nonadaptive automation, the tasks per-formed by an adaptive system, or the level of automation (LOA), are not fixed but may change in real time. Changing LOA could be handled by the human operator in an explicit task allocation loop, which would add the task LOA management to the workload of the human operator. In contrast, LOA management could also be handled by an automated decision module in an implicit loop (e.g., Tattersall & Fairclough, 2003). Ultimately, a system is envisioned that supports and interacts with the human operator in a way that reflects social relationships. For instance, it could resemble a team member or subordinate (e.g., Mulder, de Waard, Hoogenboom, Quispel, & Stuiver, 2008). In this kind of human-machine interaction, the machine has the authority to initiate changes in the number of automated subtasks, the LOA per subtask, and the type of support (e.g., Scerbo, 2001). This capability requires real-time assessment by the system to decide when support is needed.

When designing an adaptive system, one needs to establish what defines the necessity for adapting task automation. As such, several broad strategies have been suggested (e.g., Morrison & Gluckman, 1994; Parasuraman, Bahri, Deaton, Morrison, & Barnes, 1992). For example, aid could be triggered by critical events in the environment that are not directly linked to user behaviour. In contrast, the operator's psychophysiological signals could also be used to trigger task support. Another user-adaptive approach is to monitor task performance and change LOA to prevent unacceptable performance degradation (e.g., Rouse, 1988). It has also been suggested that the LOA be coupled to mental workload (MWL), here defined as "the reaction to demand; the proportion of the capacity that is allocated for task performance" (de Waard, 1996), which would ideally create a system in which the operator's mental workload remains relatively stable (e.g., Parasuraman et al., 1992).

Stabilizing mental workload, by preventing long periods of both overload and underload, has great advantages for what we think should be the ultimate purpose of an adaptive system in safety-critical tasks: to protect the operator and other humans in contact with the system. According to the operator functional state framework, for example, operating with high MWL, even if performance levels are still acceptable at a particular moment, might warrant a change of automation since acceptable performance is increasingly difficult or even impossible to sustain in demanding working conditions (e.g., Hockey, 1997, 2003). These considerations imply that MWL assessment could be used to decrease the probability of passing thresholds for unacceptable performance levels.

Although MWL assessment could be used as a trigger for task support, some hysteresis should be present in the system (Hoogeboom & Mulder, 2004) in that not all MWL changes should result in a modified support level. Dead bands will likely be necessary to prevent excessive, rapid switching between levels of support. Also, the effectiveness of an adaptive system will partly depend on the operator's long-term acceptance of the adaptive system. Therefore, operator preferences in timing might temporarily overrule triggers. Last, behavioural adaptations when operating an adaptive support system may become an issue. Human operators themselves control how much effort they will invest and may find new ways to perform their task that were not anticipated by the system's designers, which may potentially result in reductions or even reversals of potential safety gains (e.g., de Waard et al., 1999; Lee, 2008; Parasuraman & Riley, 1997). For instance, operators might try to maximize support levels, even if there is no need to reduce effort investment (Hoogeboom & Mulder, 2004).

Although setting the adaptive properties of an automated system is a major design requirement, another challenge is deciding how to support the human operator. Although available support types depend greatly on the task, some general concepts are available in the literature. The simple binary approach of manual versus fully automated systems was followed by a graded classification of LOA, including levels such as the computer "allows the human a restricted time to veto before automatic execution" and "the computer offers a complete set of decision/action alternatives" (Parasuraman, Sheridan, & Wickens, 2000, p. 287). Moreover, LOA may also be viewed as a unidimensional continuum of automation degrees (e.g., Parasuraman, Sheridan, & Wickens, 2000).

A conceptual shift away from a decision and action execution–oriented approach to LOA was to include the early stages of the human information processing system, sensory processing and perception or working memory, as candidates for automation (Parasuraman et al., 2000). Another scale of automation degrees was presented in Flemisch, Kelsch, Löper, Schieben, and Schindler (2008). On this scale, the term assistance is used when most of the task is performed by the human operator and could include support types that are not readily included in the concept of automation, such as providing information, warnings, and advice. Conversely, at the other end of the scale, the term automated is used when most of the task is automated.

One example of a task for which support could be provided is steering a vehicle. Lane keeping (steering or tracking) is a major, but relative easy, part of maintaining safe control of a vehicle (Parkes, 1991). However, the relative ease with which drivers generally keep in their lanes could be contrasted with accident statistics. For instance, the United Nations Economic Commission for Europe (2007) report on statistics of road traffic accidents in Europe and North America shows that about one third of the accidents involving personal injury or death may be related to inadequate lateral control.

Steering is usually carried out in a highly automated fashion, although a driver will direct attentional resources to the task whenever the situation demands it and, in doing so, switch from the control level of operation to the manoeuvring level (Michon, 1985). In demanding conditions, such as decreased visibility, narrow lanes, or when drivers are fatigued, it may be difficult to estimate lateral position (LP) adequately and therefore to maintain safety or coMfort zones (Summala, 2005; Summala, Nieminen, & Punto, 1996). In these instances, a driver may benefit from more accurate and reliable LP information. Providing accurate and reliable information would be an example of assisting the input stages of the human processing system (Parasuraman et al., 2000) and would constitute a relatively low degree of automation (Flemisch et al., 2008).

However, using raw (not averaged) data to trigger support, such as done by traditional lane-departure warning systems, may make a support system too reactive. Each driver displays naturally occurring and unsystematic driving behaviour variations, which may be a primary reason for drivers to maintain safety margins (Brehmer, 1990; Ranney, 1994). A single (small) driving error does not necessarily imply that the driver needs support immediately, but frequently occurring minor errors is a stronger indication that driver support is needed. Hence, time intervals may be used to create “trigger variables.” In the case of the steering task, frequent or long periods of driving near the driving lane delineation could be an indication that not enough effort is being invested in keeping the car in the centre of the lane. Another indication of inadequate effort investment would be lane departures, since these actions may set a vehicle in the direct path of oncoming traffic or on a potentially dangerous shoulder.

Last, the standard deviation of the LP (SDLP) can be used to assess changes in MWL. For example, the SDLP turned out to be a sensitive measure for overall driving capacity in experiments involving drugs and driving (O'Hanlon, Haak, Blaauw, & Riemersma, 1982; Ramaekers, Robbe, & O'Hanlon, 2000). Also, sleep deprivation and fatigue as a result of prolonged driving has been shown to increase SDLP (Anund, Kecklunda, Vadeby, Hjälmdahl, & Åkerstedt, 2008; de Waard & Brookhuis, 1991), and in a previous study, SDLP was sensitive to changes in lane width and oncoming traffic (Dijksterhuis, Brookhuis, & de Waard, 2011).

To investigate the behavioural effects of providing support related to lateral control and to assess user experiences, we conducted a study in which a support system was introduced to provide objective information related to the vehicle's LP, similar to a speedometer, by means of a head-up display (HUD). A HUD was chosen to keep the gaze directed toward the road as much as possible (e.g., Tufano, 1997). The HUD provided the driver with a continuous flow of information on lane width and LP (see Figure 4.1 for a screenshot of the HUD). To investigate support effects across a range of lateral control demands, lane width and oncoming traffic density were varied. In addition, an adaptive driver support condition was created by (de) activating the HUD automatically when exceeding lateral control performance thresholds.


**Figure 4.1.** Lane position information projected on the windshield of the simulator car

## 4.2 Method

### 4.2.1 Participants

We recruited 32 participants through poster announcements and paid them €20 for participating. However, 1 participant was excluded because of insufficient driving experience, leaving 31 participants (26 male). Ages ranged from 23 to 44 years (M= 26.1, SD= 4.4), and participants had held their full driving license for 5 to 24 years (M= 7.5, SD= 3.6). The self-reported total mileage ranged from 12,500 to 1,200,000 km (median = 40,000), and the annual mileage ranged from 2,500 to 50,000 km (median = 6,000). Finally, none of the participants reported using prescribed drugs that might affect driving behaviour.

### 4.2.2 Simulator and driving environment

The study was conducted using an ST Software driving simulator consisting of a fixed-base vehicle mock-up with functional steering wheel, indicators, and pedals. The simulator was surrounded by three 32-in. diagonal screens, with each screen providing a 70° view of the driving environment. A detailed description of the driving simulator used can be found in Van Winsum and Van Wolffelaar (1993).

For the experiment, a route was prepared consisting of a two-lane road winding through mainly rural scenery, divided into four main sections of uninterrupted road, stretching out for 9.1 km on average (about 7 min of driving time) and separated by small villages. Roads in each section consisted mainly of easy curves (about 80%) with a constant radius of 380 m and ranging in length from 120 m to 800 m. The road surface was marked on the edges by a continuous line (20 cm) and in the centre by a discontinuous line (15 cm), and outside the edges, a soft shoulder was present. The participants drove the simulated car (width = 1.60 m) through the entire route twice, bringing the driven distance up to approximately 80 km. On the experimental sections, outside the villages, speed was controlled by the simulator and set to 80 km/h to prevent potential compensatory speed reactions to the experimental manipulations.

### 4.2.3 Support triggers

The vehicle's LP, defined as the distance between the centre of the participant's car and the middle of the (right-hand) driving lane, was sampled at 10 Hz during the entire experiment. Sampled LP values were further processed into three support trigger variables: (a) the proportion of driving time in the near-edge zones, (b) the proportion of driving time in the over-edge zones, and (c) SDLP (see Figure 4.2).The near-edge zones were defined as strips 12.5 cm wide toward each lane edge and are illustrated by the grey areas in Figure 4.2a. Driving in this zone with any part of the vehicle counted as driving near the lane's edges, except if the car also entered the over-edge zone (see Figure 4.2b). Each second, the support algorithm counted the number of LP samples values inside these near-edge zones for the preceding 30 s and divided it by the total number of samples during this period (300 samples). Driving in the near-edge zones for more than 7.5 s (25%) triggered the HUD. For the second trigger variable, driving in the over-edge zones, the threshold was set to 3 s (10%). Finally, all 300 LP values were used to calculate SDLP, which triggered the HUD when more than 22 cm. During the adaptive support conditions, exceeding any of these thresholds activated the HUD, and support was deactivated. when all trigger variables were below their threshold value. In addition, to prevent a high-frequency on-off switch, a minimum time delay of 10 s was set as a dead-band.

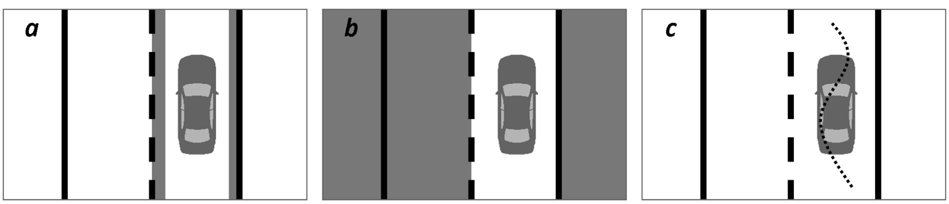



**Figure 4.2.** Abstract representation of the trigger variables. The grey areas in Figure 4.2a represent the near-edge zones toward the left and right lane edges, the grey areas in Figure 4.2b represent the over-edge zones on the left and right of the driving lane, and Figure 4.2c symbolizes the standard deviation of the lateral position.

Driving near the lane's edges, driving over the lane's edges, and weaving within the lane were assumed to be an indication of reduced lateral control, increasing the likelihood of an accident and therefore a reason to provide lateral control support. However, for a working support system, the exact width of the near-edge zone, the time length of the dead-band, and the time window used to process LP data needed to be determined in addition to the threshold values. For SDLP, Brookhuis, de Waard, and Fairclough (2003) proposed an absolute deteriorated driving criterion of 25 cm. Such clear-cut indication of threshold values for the other trigger variables do not exist in literature. During a pilot study (n= 8), a number of calibrations were tested. However, given the large number of possible calibrations, the final values were based on the rather intuitive feeling of "rightfulness." The support system was therefore calibrated in a way that most participants reported that their lateral control had worsened in the period before the support system started supplying information on LP.

### 4.2.4 Design and Procedure

Each participant completed all conditions and was instructed to drive as he or she would normally do and to follow automated auditory instructions when driving. The order of the conditions was balanced according to the Latin square method, except for the low- and high-traffic-density conditions, which alternated. Approximately 6 min of low-density oncoming traffic was always directly followed by a relatively short period of 65 s during which oncoming traffic was intensified. Hereafter, approximately 300 m before entering a village, the longitudinal control was returned to the driver. In each village, the participants received the instruction to pull over and park the car, followed by a 2-min break during which they were requested to complete mental effort rating scales.

The effects of three support modes were compared. On each road section, the HUD was either turned off (the no-support mode), continuously activated (the nonadaptive-support mode), or triggered when a threshold value was exceeded (the adaptive-support mode). In addition, lane width was either 2.25 m or 3.00 m for each road section. During the last 1,300 m of each road section, the average interval between oncoming vehicles decreased from 10 s to 1.5 s. No traffic was present on the same (driving) lane as the participants'. This resulted in a within-subject design consisting of three repeated measures factors: support mode (3), lane width (2), and oncoming traffic density (2).

Before the actual experiment started, participants were asked to read information related to the support system; this information included a full and detailed user's manual of the HUD and the exact trigger-related properties of the adaptive support mode. The main reason for providing this level of information beforehand was to make sure that the participants would understand the support system and avoid "automation surprise" (e.g., Prinzel, 2002). Furthermore, it was stressed that any HUD activation as a result of driving behaviour did not indicate bad steering behaviour but was meant to provide LP information. After providing informed consent, the participants were given the opportunity to get used to the simulator car and all support modes. The Ethical Committee of the Psychology Department of the University of Groningen approved the study.

### 4.2.5 User experience Questionnaire

After the participants were informed about the functionality of the HUD in both support modes, participants were asked to complete a technology acceptance questionnaire that consist of nine 5-point rating items that load on two factors, usefulness and satisfaction, to compare both support modes (Van der Laan, Heino, & de Waard, 1997). In addition, these expectation scores were compared with the scores given at the end of the experiment after the drivers had experienced both support types.

To detect possible differences in the support usage among the participants, they were asked to indicate in what way they had mainly used the HUD directly after the experiment. Three standard options were given: (a) "I used the lateral position information provided by the HUD," (b) "I used the HUD activation as a signal to improve lateral position behaviour," and (c) "I ignored the HUD as much as possible." In addition, the participants could tick a fourth option, "Different, namely . . . ," and provide their own answer.

During each driving break, the participants were asked to mark the Rating Scale Mental Effort (RSME; Zijlstra, 1993) for both the first part (low-density oncoming traffic) and the last part of the drive (high-density oncoming traffic). The RSME is a single-dimension rating scale ranging from 0 to 150 that can be used by the participants to rate experienced effort. In addition, several effort indications (calibrated anchor points) are visible alongside the scale that may further guide the participant's rating. Indications start with absolutely no effort (RSME score of 2) and end with extreme effort (RSME score of 112).

### 4.2.6 Analyses

We analysed the data using the General Linear Model Repeated Measures test of SPSS. Repeated-measures MANOVAs were run on the system acceptance scores (usefulness and satisfaction) and the lateral control variables. We tested LP, SDLP, and time in the near- and over-edge zones using the mean value for all time windows within an experimental condition. ANOVAs were run on the RSME scores. Alpha was set to 5% for all tests, and to com-pare individual support modes, Bonferroni post hoc tests were used when appropriate.

## 4.3 Results

### 4.3.1 User experiences

To begin with, there were differences among participants in the way the HUD was mainly used. Of all the participants, 39% indicated that they ignored the HUD as much as possible. Furthermore, within the group of participants who did not ignore the HUD, 79% used the HUD primarily as a warning signal, whereas 16% reported to have used it as a source of LP information.

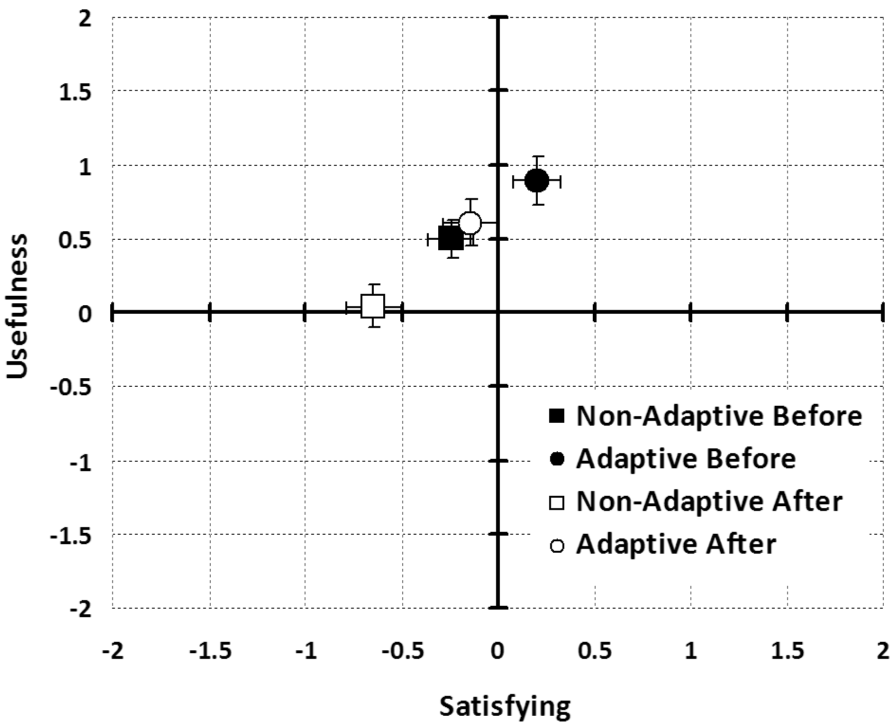



**Figure 4.3.** System acceptance scores before and after the experiment for both support modes. Error bars represent the standard error.

The subjective scores for both the usefulness and the satisfaction dimension (both ranging from –2 to +2) are shown in Figure 4.3. When looking at these two dimensions in Figure 4.3, one can see that the usefulness scores deviate most from the neutral, zero line and that adaptive support prior to the experiment received the highest score. On the satisfaction dimension, three out of four scores are within 0.5 points of the zero line. Despite these small ranges, both usefulness and satisfaction show two main effects (support type, and before and after testing). Overall, adaptive support was rated higher than the nonadaptive support on both usefulness and satisfaction dimensions: usefulness, $F(1, 29) = 9.91$, $p< .01$, $\eta_p^2= .26$; satisfaction, $F(1, 29) = 16.02$, $p< .001$, $\eta_p^2= .36$. Also, the participants' expectation scores on both dimensions were higher than were the scores after experiencing either support type: usefulness, $F(1, 29) = 10.79$, $p< .01$, $\eta_p^2= .27$; satisfaction, $F(1, 29) = 14.34$, $p< .01$, $\eta_p^2= .33$.

The mental effort ratings are shown in Figure 4.4. Mean mental effort ranges from just over the almost no effort mark for wide lane, no support, and low-density traffic to just below the some effort mark for narrow lane, no support, and high-density traffic. Several effects on mental effort were present in the data. To begin with, the wide lane conditions were rated significantly lower than the narrow lane conditions (19.9 vs. 27.6 points). Also, low-density oncoming traffic was rated lower than the high-density conditions (18.3 vs. 29.1 points). However, the differences between the support modes did not result in a significant effect (10.5, 11.9, and 11.1 points for no support, nonadaptive support, and adaptive support, respectively). Finally, one interaction effect is present. When participants were driving on a wide lane, an increase of traffic density resulted in an increase of 7.6 RSME points. When they were driving on a narrow lane, this increase was 14.0 points (see Figure 4.4 and Table 4.1).

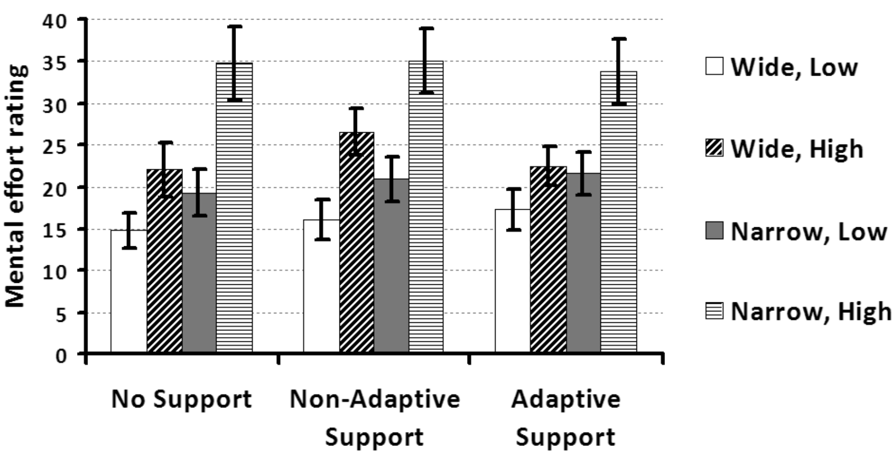



**Figure 4.4.** Reported mental effort for all support modes. Wide = wide lane (3.00 m); narrow = narrow lane (2.25 m); low = low-density oncoming traffic; high = high-density oncoming traffic. Error bars represent the standard error. Maximum score for mental effort is 150.

**Table 4.1.** Univariate Test Results for Subjective Effort (Figure 4.4). Support = support mode; width = lane width; traffic = oncoming traffic density. Significant effects ($p< .05$) are shown in **bold**. Degrees of freedom are Greenhouse-Geiser corrected when Mauchly's test showed violation of sphericity.

| *Effect* | F(df1,df2) | p | $\eta_p^2$ |
|---|---|---|---|
| Support (S) | <1(2,58) | ns | 0.005 |
| **Width (W)** | **11.58(1,29)** | **0.002** | **0.285** |
| **Traffic (T)** | **26.63(1,29)** | **<0.001** | **0.479** |
| S × W | <1 (2,58) | ns | 0.007 |
| S × T | 1.46(2,58) | ns | 0.048 |
| T × W | 11.24(1,29) | 0.002 | 0.279 |
| S × W × T | **<1 (2,58)** | **ns** | **0.033** |

### 4.3.2 Performance Measures

For LP, defined as the distance between the centre of the participant's car and the middle of the (right-hand) driving lane, main effects for all factors were found (see Figure 4.5a and Table 4.2). Driving on a narrow lane and exposure to high-density oncoming traffic were associated with an LP toward the shoulder. In addition, these effects strengthened each other, creating a significant interaction. Also, support mode affected LP. Bonferroni post hoc comparisons did not reveal any significant differences, although the largest contributor to the main effect is the difference between the adaptive and the nonadaptive support mode (mean difference = 3 cm, $p = .051$).

As can be seen in Figure 4.5b, driving on narrow lanes decreased SDLP values compared with driving on wide lanes. Also, support mode yielded a main effect. A closer look revealed that SDLP was lower for the adaptive support mode than for the no-support mode (mean difference = 2 cm, $p = .027$). Increases in traffic density did not change SDLP.

The time participants drove with any part of the vehicle within 12.5 cm of either edge of the driving lane (but not over the edge) was not affected by support mode (see Figure 4.6a and Table 4.2). However, participants did spend more time inside the near-edge zone when on narrow lanes and during high oncoming traffic density.

The last vehicle parameter refers to the time participants drove with any part of the car over the driving lane edges (Figure 4.6b), and main effects were found for all factors. An increase in time spent over the driving lanes was associated with both driving on a narrow lane and high-density oncoming traffic. However, these factors seemed to reinforce each other, resulting in an interaction. Last, a difference was found between support modes. Bonferroni post hoc analysis revealed a significant difference between adaptive support and nonadaptive support (mean difference = 3%, p = .017).

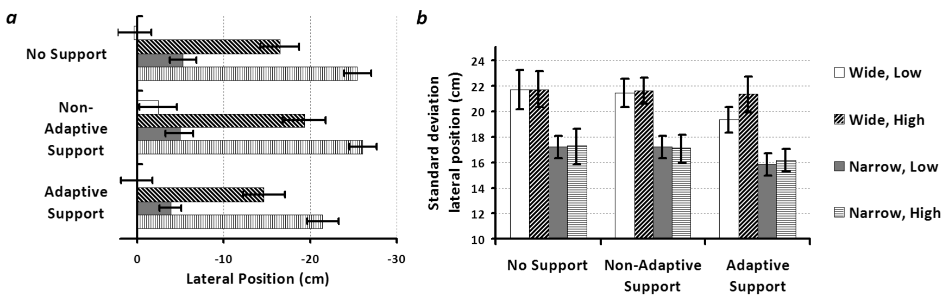



**Figure 4.5.** Mean (a) and standard deviation (b) of the lateral position for all support modes. Lateral position values represent the middle of the car (car width = 1.60 m) in relation to the middle of the right (driving) lane. Negative numbers indicate a position to the right of the lane middle and vice versa. Wide = wide lane (3.00 m); narrow = narrow lane (2.25 m); low = low-density oncoming traffic; high = high-density oncoming traffic. Error bars represent the standard error.

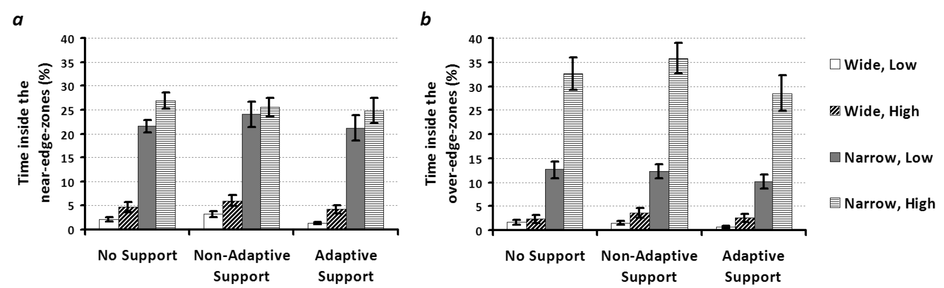



**Figure 4.6.** Percentage of the time spent inside the near-edge zone for wide lane (a) and over-edge zone (b) for all support modes. Wide = wide lane (3.00 m); narrow = narrow lane (2.25 m); low = low-density oncoming traffic; high = high-density oncoming traffic. Error bars represent the standard error.

**Table 4.2.** Univariate Test Results for Vehicle Parameters (Figures 4.5a, 4.5b, 4.6a, 4.6b). Support = support mode; width = lane width; traffic = oncoming traffic density; LP = lateral position; SDLP = standard deviation lateral position. Significant effects ($p < .05$) are shown in **bold**. Degrees of freedom are Greenhouse-Geiser corrected when Mauchly's test showed violation of sphericity.

**Vehicle parameters**

| | Mean LP | | | SDLP | | |
|---|---|---|---|---|---|---|
| *Effect* | F(df1,df2) | p | $\eta_p^2$ | F(df1,df2) | p | $\eta_p^2$ |
| Support (S) | **3.81 (2, 56)** | **0.028** | **0.120** | **3.90(2, 56)** | **0.026** | **0.122** |
| Width (W) | **31.93(1, 28)** | **<0.001** | **0.533** | **59.59(1,28)** | **<0.001** | **0.680** |
| Traffic (T) | **183.67(1, 28)** | **<0.001** | **0.868** | <1(1, 28) | ns | 0.016 |
| S × W | <1(2, 56) | ns | 0.021 | <1(2, 56) | ns | 0.005 |
| S × T | **2.63(2, 56)** | **0.081** | **0.086** | 1.02(2, 56) | ns | 0.035 |
| T × W | **9.03(1, 28)** | **0.006** | **0.244** | <1(1, 28) | ns | 0.032 |
| S × W × T | <1(2, 56) | ns | 0.012 | <1(2, 56) | ns | 0.023 |
| | Inside 'Near-Edge-Zone' | | | Inside 'Over-Edge-Zone' | | |
| *Effect* | F(df1,df2) | p | $\eta_p^2$ | F(df1,df2) | p | $\eta_p^2$ |
| Support (S) | <1(2, 37.82) | ns | 0.033 | **5.10(2, 56)** | **0.009** | **0.154** |
| Width (W) | **218.75(1,28)** | **<0.001** | **0.887** | **146.41(1,28)** | **<0.001** | **0.839** |
| Traffic (T) | **30.40(1,28)** | **<0.001** | **0.521** | **106.42(1,28)** | **<0.001** | **0.792** |
| S × W | <1(1.5,43.0) | ns | 0.000 | **2.83(1.6,44.1)** | **0.068** | **0.092** |
| S × T | 1.16(2, 56) | ns | 0.040 | 1.14(2, 56) | ns | 0.039 |
| T × W | <1(1, 28) | ns | 0.015 | **94.91(1, 28)** | **<0.001** | **0.772** |
| S × W × T | <1(2, 56) | ns | 0.027 | <1(2, 56) | ns | 0.034 |

## 4.4 Discussion

This study was designed to investigate the driving effects of nonadaptive and adaptive driver support in a wide range of driving demands and to assess drivers' experiences with using these systems. Differences between support modes were smaller both in number and in size when compared with the effects caused by manipulations of infrastructure and traffic. Effort ratings were unaffected by support mode, although most vehicle parameters did show a difference. On closer examination, it seems that these effects were mainly caused by the adaptive support. Compared with either the nonadaptive or the no-support mode, participants drove more centrally, swerved less, and drove less on the shoulder. In addition, none of the post hoc comparisons between the nonadaptive and the no-support mode were statistically significant

Instead, there were distinct differences in the manner in which the HUD was primarily used. More than a third of the participants indicated to have ignored the HUD. This behaviour might be classified as disusing the provided assistance, which is defined as a failure to engage in automation when it could improve performance (Lee, 2008; Parasuraman & Riley, 1997). What has caused these differences can only be speculated. In addition to design specifics, use of the system may correlate with personality traits relevant to driving behaviour, such as sensation seeking (e.g., Constantinou et al., 2011). Also, participants were not asked about their driving history in terms of traffic violations, accident involvement, and so on. For future research, it would be interesting to use this kind of information to establish who is more likely to ignore useful information. In addition, we cannot precisely discern which HUD version the participants had in mind when answering this question because it was asked only once, shortly after the experiment. However, the satisfaction and the usefulness scores indicate that the adaptive HUD was the most accepted.

Within the group of participants who did not mainly ignore the HUD, 79% used HUD activation as a warning signal; we can therefore conclude that the participants mainly appreciated the warning property of the adaptive HUD and not its information provision property. Although the participants did not use adaptive support as intended by the system's designers, it seemed to help driving performance. It could be argued that using naive participants might have prevented usage of the system in this way. However, this scenario is unlikely. If this system had been tested across a longer period, naive drivers eventually would have learned when support is activated.

When comparing the current system with previously investigated adaptive systems, one can find several differences and similarities. Adaptive automation is usually applied to (modified versions of) complex tasks, such as the Multi-Attribute Task (e.g., Pope, Bogart, & Bartolome, 1995), the Cabin Air Management task (e.g., Ting et al., 2010), the Multitask (e.g., Kaber & Endsley, 2004), or a simulated helicopter cockpit task (Miller & Hannen, 1999). Adequate performance of these tasks requires a lot of participant training or trained professionals. In contrast, a driving simulator used for the current study closely resembles a task environment highly familiar to most people. Other comparisons can be made with regard to the aid itself. A focus of the current study was aiding the input functions of the human information processing system (Parasuraman et al., 2000), which it has in common with the Cockpit Information Manager (CIM; Miller & Hannen, 1999), even though the LOA in the current study must be considered lower since it did not involve organizing the data in any way.

In other studies, the output functions (decision making and response selection) are supported. For example, Ting et al. (2010) developed an adaptive aid that could control the regulation of crucial air parameters. Also, compensatory tracking, as used by Pope et al. (1995), was either fully automated or fully manual. Apropos, given the focus on input functions and a relatively low LOA, the current system could be referred to as assistance rather than automation on the continuum of automation degrees presented in Flemisch et al. (2008).

Another relevant feature of the adaptive system described in the current article is the use of task performance to trigger aid, even though many systems were designed with a different approach in mind (a classification is given in Parasuraman et al., 1992). The CIM, for example, uses context information and crew actions to determine current and near-future tasks to adjust information provision. Another alternative approach was implemented by Ting et al. (2010), who combined electrocardiogram and electroencephalogram data to predict error performance, which was subsequently used to activate task support. The aforementioned comparisons represent just a selection of dimensions along which comparisons can be made. Also, more systems could have been mentioned. Regretfully, a more thorough classification of all published adaptive automated systems is beyond the scope of the current article.

A HUD was chosen as the information carrier because it increases gaze time toward the roadway environment compared with head-down displays. However, several problems with using a HUD have been mentioned in the literature. For example, the HUD's focal distance can influence perception time of both the information on the HUD and the roadway environment (Gish & Staplin, 1995; Tufano, 1997). Also, cluttering of the visual field, superimposing of salient environmental information, and poor detection of unexpected peripheral events as a result of attentional capture can be problematic (e.g., Horrey & Wickens, 2004; Weintraub & Ensing, 1992).These problems were probably less relevant for the present study. To start with, driving in a rural environment and the HUD position (see Figure 1) minimized the superimposing of oncoming traffic. Also, the visual workload demanded by the driving situation can be considered quite low, making it unlikely that adding the HUD overloaded visual attentional resources. Last, peripheral detection was not an issue in the simulator, since salient peripheral objects were not present. However, the issues associated with using HUDs are extremely relevant in a real-world application and might be overcome by using what has been called "conformal imagery," a form of augmented reality (Fadden, Ververs, & Wickens, 2001).

A limitation of the study may be the calibration of the adaptive support system. For the adaptive system, several parameters needed to be set, such as the threshold values, and extensive testing of the complete range of calibrations would have been a daunting enterprise. To solve this problem, we specifically aimed at avoiding a calibration that would make the adaptive system appear unfair or unjust as evaluated by the drivers during a pilot study. System calibration was partly based on trial and error but ultimately on the feeling of rightfulness by the pilot test participants. Using a different calibration for the adaptive support mode may have produced different experimental results. Speculatively, using stricter LP-related thresholds (driving time in the near and over edge zones) could have resulted in more HUD activations and therefore in driving more centrally. However, it is not clear how (de)activation frequency would interact with the willingness to use the provided information; this question may be a topic for future research.

Another limitation is the implementation of the high-density oncoming traffic conditions, which were shorter than the low-density conditions and occurred in a fixed, alternating order. Alternating the traffic density levels was done to prevent interference from the high-density condition in the quiet condition. It is unclear whether a different design would have produced different results. However, in hindsight, a balanced order and equal duration of these conditions would have been preferred to decrease the likelihood of order effects and to exclude the possibility that condition length confounded the large effects of increasing oncoming traffic density.

Last, only lane-keeping performance was used to assess MWL changes even though a change in performance does not always equate to a change in MWL. Changes in performance levels may also be the result of changed task demands, whereby performance might degrade even if the MWL level is sustained. Also, a depletion of energetic and cognitive resources, despite intentions of good performance, might prevent further increase in MWL (e.g., Hockey, 1997). Considering the flexibility displayed in human behaviour, one may conclude that it is unlikely that a reliable single measure for MWL exists and that several sources (subjective, performance, environment, and psychophysiology) are required for deriving a more reliable index (Brookhuis & de Waard, 2000; de Waard, 1996; Gaillard & Kramer, 2000; Hoogeboom & Mulder, 2004).

Using several sources, such as physiological triggers, for example, would have potentially increased the range of MWL detection to beyond the point at which performance cannot be increased anymore. However, as the intention was to assess changes in MWL, rather than absolute levels, it is carefully assumed that within a highly controlled experimental setting, a performance change will likely be caused by a change in resource allocation in service of the steering task and, therefore, a change in MWL.

### 4.4.1 Conclusion

Across a range of driving demands, driving performance was somewhat improved when drivers were exposed to an adaptive support system that provided feedback on LP. This was likely caused by unexpected use of the system. Participants indicated to have mainly used the moment of support activation as a warning signal rather than using the information provided on the HUD itself. In this way, the system not only supported the perception phase but supported the decision-making phase as well. It seems unlikely that an adaptive support system can be designed that will work only for the pre-decision stages; indicating that all perceived properties of an adaptive system are likely to be utilized by the human operator.

### Acknowledgments

The authors would like to thank Christian Kroiß, especially for his assistance in creating the adaptive head-up display software module, and Ben Lewis-Evans, for proofreading the manuscript. This research was supported by the Commission of the European Communion, Project No. 215893 REFLECT, Responsive Flexible Collaborating Ambient

# Chapter

# 5

# The potential of music selection for adaptive driver support

Chapter 5 is based on van der Zwaag, M. D., Dijksterhuis, C., de Waard, D., Mulder, B. L., Westerink, J. H., & Brookhuis, K. A. (2012). The influence of music on mood and performance while driving. Ergonomics, 55(1), 12-22. DOI: 10.1080/00140139.2011.638403

**Abstract**

An adaptive driver assistance system may potentially influence driving behaviour through changing the music that the driver is listening to. Before such a system can be tested, the influence of listening to music while driving need to be examined in more detail. In the current study, drivers in a simulator listened to individually negatively and positively rated music while driving on rural roads. High demanding driving conditions were implemented by narrowing the lane's width. Results indicate that when driving on a narrow lane, swerving decreased and subjective effort ratings increased, indicating an increase in mental workload. Effects of listening to music on driving behaviour were limited to a marginal effect of music type: driving speed decreased as a result of listening to positively rated music compared to the no-music condition. In addition, breathing rate slowed down when listening to negatively rated music compared to the no-music condition. These results indicate that a driver assistance system that automatically selects music valence, is unlikely to immediately influence driving behaviour in these driving conditions. On the other hand, decreased breathing rates signal a relaxed state, which could be used to counter unwanted affective states.

## 5.1 Introduction

In Western society, music listening has become a frequent activity in the background of almost any activity (DeNora, 2000; North & Hargreaves, 2008). Music researchers have now started to focus on music listening in these specific everyday life situations to improve the understanding of how music can influence behaviour (DeNora, 2003; Juslin & Sloboda, 2010). Driving is one of the most popular music listening situational contexts. While driving, people listen to music to attain enjoyment or to feel engaged when driving in solitude (DeNora, 2000; Walsh, 2010). It is also suggested that music listening distracts from driving and can therefore influence safety (Brodsky, 2001). Although the impact of music on driving performance has been given some attention (Dibben & Williamson, 2007), its impact on physiological measures has not. Neither has a distinction been made between the respective impacts of the specific types of music such as positive and negative valenced music. In the current article, these relationships between music valence, physiological measures, and driving performance are studied.

### 5.1.1 Music and physiological measures

In the literature, several effects of listening to music can be found. To start with, fast tempo music has consistently shown to increase arousal levels compared to slow tempo music (e.g., Krumhansl, 1997; Van der Zwaag et al., 2011). The majority of these studies have found that arousing music increases heart rate compared to low arousing music (De Jong et al., 1973; Knight and Rickard, 2001). Still, others have found that any music, both low and high arousing, increases heart rate (Krumhansl, 1997; Iwanaga & Moroki, 1999; Rickard, 2004). Respiration rate is found to increase in high arousing music compared to relaxing music as well (Iwanaga and Moroki, 1999; Krumhansl, 1997; Nyklíček et al., 1997; Gomez & Danuser, 2004). Again, in other studies no difference in respiration rate while listening to different types of music was found (e.g., Davis, 1992). Hence, inconsistent results are found on heart rate and respiration rate responses to music listening.

Several explanations can be given for the inconsistent results found for the influence of music on physiological measures. A first explanation comes from the fact that the studied physiological measures are affected by regulatory effects in the autonomic nervous system, which is primarily responsible for keeping homeostasis (Cacioppo et al., 2000). As a result, physiological responses are not solely influenced by music listening (e.g., through an emotional response) but additionally, via physical activity, cognitive demand, and other psychological constructs (Cacioppo et al., 2000; Van den Broek & Westerink, 2009). Hence, the situational context should be taken into account in interpreting physiological responses to music listening. A second explanation can be found in the fact that most studies in music research differ to a great extent on important methodological aspects, such as the song selection method and the duration of the music presentation. For the study of physiological responses to music, awareness of these methodological aspects is important, as is the perspective to always describe the impact of music in relation with personal and situational context (Blacking, 1973; Saarikallio & Erkkilä, 2007; North & Hargreaves, 2008; Sloboda & Juslin, 2010).

### 5.1.2 Music while driving

Music can be beneficial while driving as, for example, the arousal hypothesis predicts that in cases of boredom and drowsiness music can lead to a more optimal arousal level for driving (North & Hargreaves, 2008; Shek & Schubert, 2009). However, following the distraction hypothesis, music can also take attention away from the driver (Shek & Schubert, 2009). This distracting effect of music on driving can be disadvantageous when it decreases safety in case high arousing music is played during high demanding road situations (Dibben & Williamson, 2007). However, Wiesenthal et al. (2000) showed that one's favourite music alleviates stress during high congestion drives. These authors found higher stress levels when comparing no music to favourite music during high congestion drives. Furthermore, it is shown that driver aggression can be tempered with favourite music compared to no music in high demanding rides (Wiesenthal et al., 2003).

Explanations for the effects of music listening while performing a concurrent task, such as driving, often focus on processing capacity in service of the primary task (North and Hargreaves, 1999; Dalton and Behm, 2007; Pêcher et al., 2009), assuming that listening to music may be arousing and requires mental resources. Following the information-distraction approach, music adds additional irrelevant stimuli to a task which leads to increased cognitive load and thus can impact task performance (Konecni, 1982; North & Hargreaves, 1999; Recarte & Nunes, 2000). Consequently, the more attention a particular type of music requires the more it competes for processing resources with the primary task of, for example, driving. To illustrate, North & Hargreaves (1999) manipulated cognitive load of participants by exposing them to low or high arousing music by varying tempo and volume in a driving game. They found that high arousing music resulted in worse racing performance defined as slower lap times, while the quickest lap times were recorded when listening to low arousing music. Interestingly, they also found a connection between cognitive load and music liking, and concluded that competition for processing resources caused participants to dislike music. Pêcher et al. (2009) mentioned that post-experiment interviews revealed that drivers found happy music the most disturbing and, combined with behavioural data, took this as support to conclude that listening to happy music resulted in deteriorated driving performance.

The impact of in-vehicle music listening on driving speed may depend on the road situational context. Reducing speed is found to be used as a compensatory reaction when faced with high load situations. Namely, it enables the driver to maintain safety margins by decreasing required reaction times (Summala, 2005). Furthermore, as mentioned above, it can be expected that drivers allocate more attention, and thus mental resources, to positive music, which could result in detrimental effects on vehicular control or in a compensatory reaction such as slowing down.

Because task performance and music listening might compete for the same mental resources, the impact of musically evoked cognitive demand on performance might be dependent on the cognitive demand of a concurrent task (Konecni, 1982; North & Hargreaves, 1999). In low demand driving situations there is less competition for attentional space. Hence, it is likely that mental resources can more easily be divided between listening to music and driving as the limits of mental resources are not reached. Therefore, listening to music should have a low impact on driving performance in these low-demand situations. As lane width is known to influence driver's workload, this variable could be used to manipulate primary task demands when studying the relation between listening to positively and negatively rated music. Results reported in the literature show that when driving on narrow lanes, less manoeuvring space is available for the driver, and more attention is required to prevent driving errors such as drifting out of the driving lane and to maintain personal safety margins (de Waard et al., 1995; Dijksterhuis et al., 2011). This results in smaller deviations from the driver's preferred lateral position (LP) on the road (de Waard et al., 1995; Dijksterhuis et al., 2011) and a compensatory speed reduction (Godley et al., 2004).

### 5.1.3 Expectations

To start with, while driving, we expect an increase in respiration rates and heart rate during high (narrow lane width) compared to low (wide lane width) demanding drives. In accordance with Dijksterhuis et al. (2011), less swerving (i.e., reduced variation in LP) is expected on narrow lanes. Furthermore, width reduction could lower driving speed to compensate for the higher amount of resources allocated in the more demanding drive (de Waard et al., 1995; Godley et al., 2004). Finally, we expect that music would influence driving performance in high demand drives, as in those conditions music competes with the limited amount of mental resources available.

## 5.2 Method

### 5.2.1 Participants

The study had been approved by the ethics committee of the Department of Psychology of the University of Groningen and informed consent was obtained from all participants. Nineteen participants, 13 men and 6 women, were paid 45 Euros for participating. Age ranged from 22 to 44 years (mean=27.5; SD=5.2) and participants had held their driving licence for 4 to 22 years (mean=8.8; SD=4.9). Self-reported total mileage ranged from 6000 to 700,000 km (median =45,000; inter-quartile range (IQR)=77,500 km) and current yearly mileage ranged from 1500 to 60,000 km (median=7000; IQR=5000).

### 5.2.2 Design

Three music conditions were included: positive music, negative music and no music. In addition, two levels of lane width (wide 3.00 m or narrow 2.50 m) were created in the driving simulator, corresponding to low and high demand drives, respectively (Dijksterhuis et al., 2011). Participants completed four sessions on separate days: one introduction session and three experimental sessions. In each experimental session, one music level was presented and both lane widths were used. This resulted in a within-subject design including two repeated-measures factors: music (3) and lane width (2). The order of both the music and lane width factors were counterbalanced among participants.

### 5.2.3 Music stimuli selection

Because music preference is highly personal (e.g., Hargreaves & North, 2010), songs used as stimuli were selected individually. To do so, participants completed an introductory session prior to the experimental sessions. In this session, participants rated 60 songs on perceived valence and energy levels on 7-point Likert scales. The participants did not have to listen to the entire song but were encouraged to sample each song for a few moments and at a few locations within the song to get a good impression of the song. The 60 included songs were selected to have a large range of valence and energy values and were selected from a database containing 1800 songs in total. The songs were selected based on energy and valence labels which were acquired by automatic classification of music characteristics (Skowronek et al., 2006, 2007). The order of the song presentation was randomised over participants.

After participants had finished the ratings, nine songs were selected per participant and per music condition (positive/negative) in such a way that valence ratings differed as much as possible between the positive and negative songs while keeping energy ratings as average as possible. Subsequently, three of the selected songs were used for the music mood induction, three songs for the high demand drive, and three songs for the low demand drive. The duration of the three songs was adjusted, using Audacity (Version 1.2.4), to 8 min, keeping the average duration of each song about equal. This was done by cutting the song to about 2.45 min and fading out the new ending of the song.

To check the selected song stimuli on their valence and energy ratings, a repeated-measures ANOVA with music (positive/negative) as within-subject factor on the valence and energy ratings of the selected songs was conducted. Results showed a main effect of music on both energy and valence ratings: valence $F(1,17) = 231.20$, $p < 0.001$, $\eta_p^2=0.93$; energy $F(1,17) = 16.04$, $p < 0.001$, $\eta_p^2 = 0.49$. This confirmed that the selected song stimuli for the two music conditions were significantly different from each other in valence and energy. The positive songs showed higher valence and energy ratings compared to the negative songs; positive songs mean (SE) valence M = 5.8 (0.17), energy M = 4.7 (0.20), negative songs valence M = 2.3 (1. 4 ), energy M = 3.3 (0.26), on scales running from 1 to 7.

### 5.2.4 Simulator and driving conditions

The study was conducted using a STSoftware© driving simulator. This simulator consists of a fixed-base vehicle mock up with functional steering wheel, indicators, and pedals. The simulator was surrounded by three 32" diagonal plasma screens. Each screen provided a 70° view, leading to a total 210° view. A detailed description of the functionality of the driving simulator used can be found in Van Winsum & Van Wolffelaar (1993).

Participants drove the simulated car (width: 1.65 m) over two sections of uninterrupted two-lane roads (2.50 m or 3.00 m wide lanes), winding through rural scenery, and separated by a small town. Roads in each section consisted mainly of easy curves (about 80%) with a constant radius of 380 m and ranging in length from 120 to 800 m. The road surface was marked on the edges by a continuous line (20 cm wide), in the centre by a broken line (15 cm), and outside the edges a soft shoulder was present. The posted speed limit during the drive was 80 km/h. In addition, a stream of oncoming traffic was introduced with a random interval gap between 2 and 6 s, resulting in 15 passing passenger cars (width: 1.75 m) per minute on average. No vehicles appeared in the participant's driving lane.

### 5.2.5 Measures

#### 5.2.5.1 Subjective ratings

Subjective mood scores of valence (ranging from unpleasant to pleasant) and energy (ranging from tired/without energy to awake/full of energy) and calmness ratings (ranging from tense to calm) were assessed using the UWIST Mood Adjective Checklist (UMACL) (Matthews et al. 1990). This UMACL contains eight unipolar items for each dimension, which start with: 'right now I am feeling. ...', and range from 1: 'not at all' to 7: 'very much'. Results of this checklist are not further reported in this thesis. However, the interested reader is referred to Van Der Zwaag et al. (2012) for a full description of the results.

The rating scale mental effort (RSME) was used to assess mental effort (Zijlstra, 1993). The RSME is a one-dimensional scale, ranging from 0 to 150, used to rate mental effort. In addition to digits, several effort indications (calibrated anchor points) are visible alongside the scale to further guide rating. Indications start with 'absolutely no effort' (RSME score of 2) and end with 'extreme effort' (RSME score of 112).

#### 5.2.5.2 Physiological measures

Physiological measures covered reactions in the cardiovascular and respiratory domain. The Portilab data recorder and its accompanying sensors were used to record these responses with a sample frequency of 250 Hz (version 1.10, Twente Medical Systems International, Oldenzaal, the Netherlands). Physiological measures were assessed continuously during the experiment. The MATLAB programming environment (2009, The Mathworks, Natick, MA) was used for the pre-processing of the respiration signals.

Cardiovascular measures were recorded using an electrocardiogram (ECG) using three Ag-AgCl electrodes, which were placed following the standard lead II placement (Stern et al., 2001). R-peaks in the ECG signal were detected automatically, after amplification and filtering of the signal (Butterworth band pass: 0.5–40 Hz). Subsequently, the distances between successive R- peaks, the interbeat intervals (IBI), were calculated.

Respiration was recorded by means of a respiration belt (Respitrace TM, Twente Medical systems). To obtain the respiration measures, noise was excluded from the raw signal and movement artifacts were reduced by a 0.005–1.0 Hz band pass filter. The amount of respiration cycles per minute indicated the respiration rate (Wientjes, 1992; Grossman & Taylor, 2007).

#### 5.2.5.3 Driving parameters

Speed and Lateral Position (LP) were sampled at 10 Hz. Lateral position is defined as the difference in metres between the centre of the participant's car and the middle of the (right hand) driving lane. Positive LP values correspond to deviations towards the left-hand shoulder and negative values correspond to deviations toward the right-hand shoulder. The sampled LP values were used to calculate mean LP and the standard deviation of LP, or swerving behaviour.

### 5.2.6 Procedure

Participants were invited four times to the driving simulator facility of the University of Groningen. During the first introductory session, the participants were informed about the experiment, signed an informed consent form, drove a 6-minute practice drive, and completed the music rating.

During the three subsequent experimental sessions, physiological sensors were attached and participants were seated in the simulator chair. Next, physiological baseline values were acquired in a habituation period during which participants watched a Coral Sea diving movie for 8 minutes (Piferi et al., 2000). Hereafter, participants filled out the UMACL. Then an 8-minute music mood induction period started in which the participants were asked to listen to the music. To remain attentive to the music, participants were asked to listen to the music carefully to be able to answer questions regarding the music after the entire experiment. During the control session in which no music was presented, participants were asked to sit and relax for 8 minutes. The participants were not informed that mood induction took place during these 8 minutes, as this could bias the results. After 8 minutes, the participants filled out the UMACL. Again, the interested reader is referred to Van Der Zwaag et al. (2012) for a full description of the UMACL results.

Next, the first simulated drive began. The participants were instructed to drive as they would normally drive. After approximately 8 minutes, participants were instructed to park the car and the music was stopped. During this break participants were asked to complete the UMACL questionnaire and the RSME scale. Next up, the second 8-minute drive started which only differed from the first drive in lane width. After completing the ride, participants filled out the UMACL and RSME again. The total duration of each experimental session including instructions and attaching and de-attaching the physiological equipment approximated 70 minutes.

### 5.2.7 Data analysis

Data were analysed using SPSS 17 for Windows (SPSS Inc., Chicago,IL) with level of significance at $p<0.05$ (two-tailed). Both the physiological and subjective data acquired during the music mood induction period were analysed to confirm that successful mood induction took place using a repeated-measures ANOVA with music (positive/negative/no music) as within-subject variable. Furthermore, the data obtained during the drives were analysed to show the effect of the music on driving using a repeated-measures ANOVA with music (positive/negative/no music) and driving demand (wide/narrow) as within-subject variables. Pairwise comparisons were Bonferroni corrected.

## 5.3 Results

### 5.3.1 Subjective ratings

To evaluate the perceived amount of mental effort (RSME scores) during the drives, a repeated-measure ANOVA of music (positive/negative/no music) with driving demand (wide/narrow) as within-subject factors was conducted on the RSME ratings. A significant effect of driving demand was found ($F(1,18) = 9.12$, $p = 0.007$, $\eta_p^2 = 0.34$), as driving on a narrow lane was perceived as more demanding ($M = 39.37$, $SE = 5.42$) compared to the wide lane ($M = 33.10$, $SE = 4.68$). No significant main effects of music type were found, nor interaction of music type with driving demand were found; music $F(2,36) = 1.96$, $p = 0.156$, $\eta_p^2 = 0.10$; music with driving demand $F(2,36) = 0.10$, $p = 0.907$, $\eta_p^2 = 0.01$.

### 5.3.2 Physiological responses

A repeated-measures MANOVA with music (positive/negative/no music) as within-subject factor was conducted for both the respiration rate and the mean IBI duration obtained during the last 3 minutes of the baseline period. Results do not show a significant main effect of music on respiration rate and on mean IBI, indicating that the baseline respiration rates and IBI durations did not differ for the different sessions; respiration rate $F(2,36) < 1$, $p = 0.540$, $\eta_p^2 = 0.04$, mean (SE) in breath/minute: positive 14.61 (0.95), negative 14.04 (0.73), no music 15.20 (0.76); IBI $F(2,36) = 1.18$, $p = 0.360$, $\eta_p^2 = 0.62$, mean (standard error) IBI duration in seconds: positive = 0.874 (0.03), negative = 0.846 (0.03), no music = 0.876 (0.03). Next, physiological reaction scores were created by subtracting the average values obtained during the last 4 minutes of the baseline period with the values obtained during the drives.

A repeated-measure ANOVA was conducted with music (positive/negative/no music) with driving demand (wide/narrow) as within subject factors on the respiration rate. Results show a main effect of music; $F(2,36) = 3.25$, $p = 0.050$, $\eta_p^2 = 0.153$. Pairwise comparisons show a significantly lower respiration rate during the negative compared to the no music condition ($p = 0.046$; see Figure 5.1) irrespective of driving demand, see also Figure 5.1. No significant effects of the driving demand or interaction effect of music with driving demand were found.

Next, a repeated measure ANOVA was conducted with music (positive/negative/no music) and driving demand (wide/narrow) as within-subject factors on the average IBI durations obtained during the drives. Results show no significant effect for music or driving demand (all $p > 0.05$) positive mean (Standard Error) = -0.017 (0.009), negative = -0.019 (0.008), no music = -0.022 (0.007), wide lane = -0.018 (0.07), narrow lane = -0.019 (0.07).

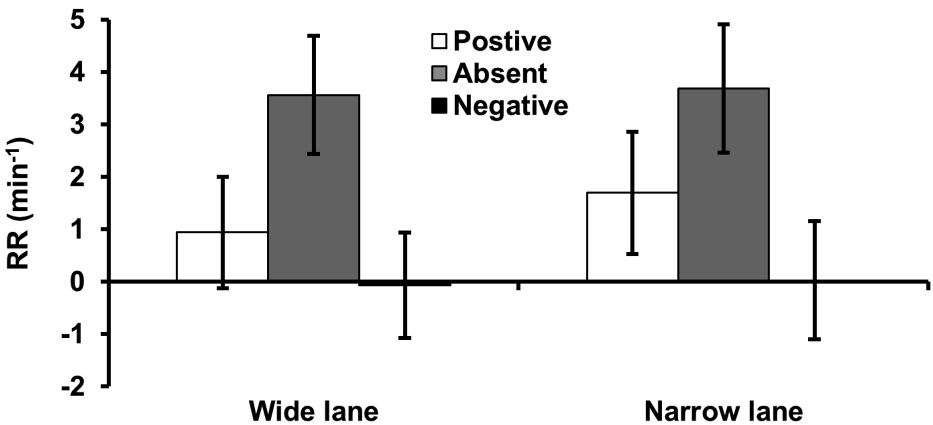



**Figure 5.1.** The average respiration rates (RR) obtained during the wide (3.00 m) and narrow (2.50 m) lane drives, relative to the last four minutes of the baseline measurement. The error bars represent the standard error

### 5.3.3 Driving performance

Separate repeated-measures analyses of music (positive/negative/no music) with driving demand (wide/narrow) as within-subject factors were conducted for the mean of the LP, the standard deviation of the lateral position (SDLP, swerving), and the speed. A marginal effect was found for mean LP ( $F(1,17) = 3.51$, $p = 0.082$, $\eta_p^2 = 0.20$) and for SDLP a significant main effect of driving demand was found ($F(1,17) = 23.80$, $p < 0.001$, $\eta_p^2 = 0.58$). During the narrow lane drive, participants drove more towards the shoulder, while the SDLP was reduced compared to the wide lane drive; see also Figures 5.2A and 5.2B. The results on speed show a marginally significant main effect of music ($F(2,13) = 3.18$, $p = 0.075$, $\eta_p^2 = 0.33$). Pairwise comparisons indicate that higher driving speeds during the no music condition compared to the positive music condition ($p = 0.023$); the average speed values are illustrated in Figure 5.2C.

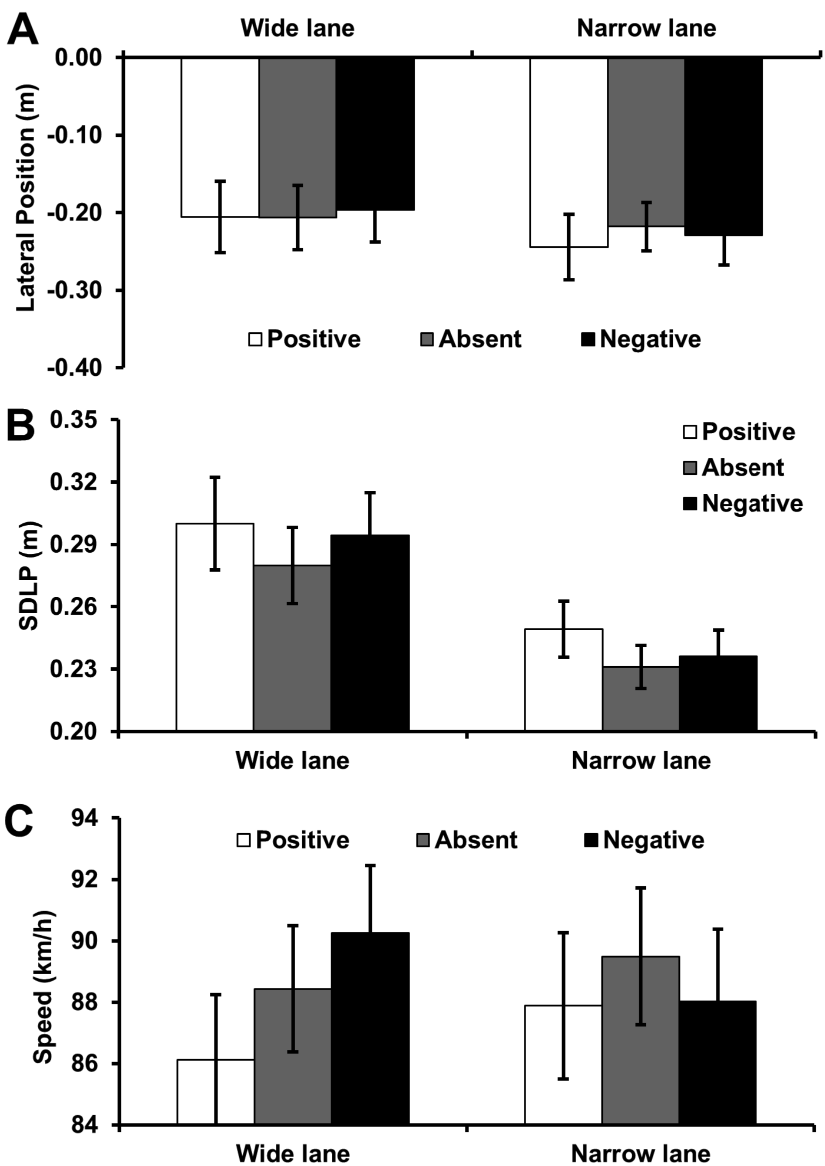



**Figure 5.2.** The average lateral position from the centreline of the road (A), the standard deviation of the lateral position; swerving (B), and driving speed (C) obtained during the wide (3.00 m) and narrow (2.50 m) lane drives. (A) Negative values indicate a lateral position towards the right hand shoulder. (A-C) The error bars represent the standard error.

## 5.4 Discussion

Music listening is a very popular side-activity while driving. However, the influence of music listening on the body state and on driving performance is not yet fully understood. Therefore, its potential for use in an adaptive driver support system is also unknown. In the current study it was investigated whether personally selected positive and negative music influences body state and driving performance. If it does, it could lead to applications that automatically select music to change driving performance.

As expected, swerving behaviour decreased and subjective effort ratings increased when driving on narrow lanes, while a compensatory speed reduction did not occur. This confirms that more effort was put into the lane keeping task to deal with decreased lateral margins and is in accordance to Summala's Multiple Monitor Theory (Summala, 2005, 2007), which states that mental load increases when less time is available to maintain safety margins. In a more general sense, these results can be seen as an indication that more effort was invested in the driving task to prevent performance degradation (Hockey, 1997, 2003) or that the level of effort was matched to the current task demands (Hancock & Warm, 1989; Matthews & Desmond, 2002).

Three interesting effects of listening to music were found. Firstly, there was an indication that positive music reduced driving speed compared to not listening to music. This result suggests that selecting positive music can be used by an adaptive system to increase driving safety. On the other hand, this effect may have been caused because drivers like listening to positive music and hence devote more attentional resources to it. The speed reduction could therefore be a compensatory reaction, which is in line with Pêcher et al. (2009). Secondly, listening to negative music compared to no music while driving resulted in lower respiration rates irrespective of driving demand. This implies that music might be used to decrease body stress of the driver as respiration rate has been linked to arousal (Boiten et al., 1994; Nyklíček et al., 1997; Ritz, 2004; Homma & Masaoka, 2008). Thirdly, in contrast to Pêcher et al. (2009), listening to music did not lead to deteriorated lateral control. Therefore, our results suggest that music listening did not affect lateral safety. This difference in results can be explained in that Pêcher et al. (2009) alternated music periods with silent periods of 1 minute each. This alternation might have distracted the drivers and therefore increased swerving behaviour while listening to music. The current results thus show a more ecological valid situation and hence more ecologically valid results.

### 5.4.1 Limitations and future research

The song stimuli varied in valence and to a lesser extend in energy levels as well. It appeared impossible to select songs that solely varied in mood valence and having equal energy levels. This implies that valence and energy ratings are not fully independent of each other in inducing mood with music. This result is in line with findings of psychobiological theory of aesthetics (Berlyne, 1971). This theory proposes that a U-shaped relationship exists between arousal and liking. That is, an average arousal level has the optimal liking level, and increasing and decreasing arousal levels would decrease liking levels (Berlyne, 1971; Hargreaves and North, 2008).

To cope with the large individual differences in music liking, individually selected song stimuli were chosen in the current study (Juslin & Sloboda, 2010). This method assured that the selected songs indeed induced the targeted moods. However, this method also resulted in stimuli that were not controlled for other music characteristics such as familiarity, or characteristics inherent to the music as tempo or mode, which could impact mood (Juslin & Sloboda, 2010; van der Zwaag et al., 2011).

One of the aims of the current study was to assess the feasibility of using music selection to regulate driving performance. For example, an adaptive driver assistance system could change music to change physiological state and thereby the driver's mental state and driving behaviour when it detects physiological or performance based indications of increased safety risks. This would create a feedback loop in which the system supports the driver by preventing suboptimal mental states such as high or low mental workload. The relatively small effect of music listening on driving indicates the unlikeliness of using music valence selection to directly manipulate workload, although this conclusion may not be true for all types of music, and may be different for other types of road, e.g., motorways. On the other hand, the marginal speed reduction as a result of listening to positively rated music seems to indicate that attentional resources may have been drawn away from the driving task. This suggests that a support system may automatically turn off music in high demanding driving situations to prevent this. Finally, effects on physiological state suggest that a support system may use music selection to help the driver to regulate bodily/affective states such as relaxedness (respiration rate), which could in turn impact driving behaviour.

### 5.4.2 Conclusion

In the current study, the influence of listening to music on body state and driving performance during high and low demand drives were investigated. Listening to negatively rated music compared to no music while driving led to decreased respiration rate and listening to positive music compared to no music leads to slower driving speed. In the present study, music did not impair driving performance nor increase subjective mental effort expenditure, which is in contrast with findings in the literature. Finally, this study indicates that automated music valence selection may be an unlikely candidate mechanism through which mental workload can be increased or decreased directly. However, it may be applied to influence the driver's physiological state, which could in turn positively impact driving behaviour.

### Acknowledgements

This work was supported by the European 7th framework REFLECT project. We would like to acknowledge Wilmer Joling, Joop Clots, and Peter Albronda, for the technical support while setting up this study.

# Chapter

# 6

# Classifying visuomotor workload from brain waves

Chapter 6 is based on Dijksterhuis C, de Waard D, Brookhuis K, Mulder B, & de Jong R (2013). Classifying visuomotor workload in a driving simulator using subject specific spatial brain patterns. Frontiers in Neuroscience. 7:149. DOI: 10.3389/fnins.2013.00149

**Abstract**

A passive Brain Computer Interface (BCI) is a system that responds to the spontaneously produced brain activity of its user and could be used to develop interactive task support. A human-machine system that could benefit from brain-based task support is the driver-car interaction system. To investigate the feasibility of such a system to detect changes in visuomotor workload, 34 drivers were exposed to several levels of driving demand in a driving simulator. Driving demand was manipulated by varying driving speed and by asking the drivers to comply to individually set lane keeping performance targets. Differences in the individual driver's workload levels were classified by applying the Common Spatial Pattern (CSP) and Fisher's linear discriminant analysis to frequency filtered electroencephalogram (EEG) data during an off line classification study. Several frequency ranges, EEG cap configurations, and condition pairs were explored. It was found that classifications were most accurate when based on high frequencies, larger electrode sets, and the frontal electrodes. Depending on these factors, classification accuracies across participants reached about 95% on average. The association between high accuracies and high frequencies suggests that part of the underlying information did not originate directly from neuronal activity. Nonetheless, average classification accuracies up to 75–80% were obtained from the lower EEG ranges that are likely to reflect neuronal activity. For a system designer, this implies that a passive BCI system may use several frequency ranges for workload classifications.

## 6.1 Introduction

In contrast to an active Brain-Computer Interface (BCI) which allows users to engage in volitional thought control of a device, several BCI researchers have proposed to advance human-computer interaction by triggering machine actions based on inferences of the user's current mental state, known as passive BCI (Cutrell & Tan, 2008; Kohlmorgen et al., 2007; Zander et al., 2010; Zander & Kothe, 2011). For example, Kohlmorgen et al. (2007) showed that it is possible to classify mental workload elicited by a secondary task mimicking cognitive processes in a real driving environment. Moreover, these classifications were then used to switch on and off a tertiary task that mimicked an interaction with the vehicle's electrical devices that in turn improved performance on the secondary task.

In the human factors and ergonomics literature, which traditionally focusses on overall system performance and safety critical tasks, the potentially detrimental effects of both mental underload and overload have been a major research topic for decades. Mental workload can be defined as a 'reaction to demand' and 'the proportion of capacity that is allocated for task performance' (de Waard, 1996). Mental underload and overload both represent compromised functional states during which a breakdown of primary task performance is more likely (e.g., Hockey, 1997, 2003; see also Brookhuis & de Waard, 2010). Preventing these hazardous functional states by maintaining mental workload or task demand within an acceptable range in real-time has been the central goal of adaptive automation since the seventies (Chu & Rouse, 1979; Hancock & Chignell, 1988; Rouse, 1988; Parasuraman et al., 1992; Kaber & Prinzel, 2006).

A large part of adaptive automation literature is devoted to determining the right moment of providing or withdrawing task support, and several types of triggers may be available to optimize performance of a human-machine system (e.g., critical events and human task performance; see Parasuraman et al., 1992). Therefore, the question arises as to what physiological measures could offer in terms of improving the overall system's performance. The most important argument for the inclusion of physiological measures in a control loop is their potential for detecting user states that would otherwise remain hidden. Human beings may exhaust themselves to protect primary task performance in demanding situations. While performance protection is important for dealing with short bursts of task demand, when exposed to longer periods of high workload, it may have affective costs such as increases in anxiety, but also compensatory performance costs, such as neglecting secondary tasks (Hockey, 1997, 2003). Since straining effort expenditure has a neurophysiological base, the ability to reliably classify workload using physiological measures could be used to offload a person, before performance effects become apparent.

Traditional research approaches might not be well suited for uncovering the underlying neurophysiological mechanisms that could be used in a support system. As pointed out by Fairclough (2009), the fundamental problem of using physiological measures is the complex relationship between user states, such as mental overload, and their associated physiological variables. Specifically, four physiology-to-state mappings can be distinguished (Cacioppo et al., 2000). In the most straightforward case, there is a unique one-to-one mapping between a physiological variable and the psychological construct. Such a unique, one-to-one mapping would be ideal for an interactive system. However a one-to-one mapping that holds true in both the laboratory and the field has to date not yet been found. A many-to-one mapping is more complicated as several signals are needed to infer a mental state. For example, heart rate, heart rate variability and blood pressure have been combined to infer mental workload (e.g., Mulder et al, 2009). In a one-to-many mapping, one physiological signal responds to a range of user states. For instance, systolic blood pressure information was found to be sensitive to excitement, frustration, and stress (Cacioppo & Gardner, 1999). Lastly, the most common finding is a many-to-many mapping where many signals are in fact sensitive to many mental states. Ultimately, in an implicit human-machine control loop, a one-to-one or a many-to-one relation is required. As briefly mentioned, another factor complicating the relationship between physiological measures and user state is lack of generalizability outside the laboratory setting where a mapping was found. Simply put, a relation between a physiological measure and a user state found in the laboratory may not hold true in a real world setting where environmental conditions are less controlled.

Furthermore, due to large individual differences in physiological responsiveness, traditional statistical tests might not be suitable to uncover relationships that are valuable for implicit machine control. Even in a repeated measures analysis of variance, where the variations due to individual differences are partly taken out of the error term, the directions of effects within the individuals need some consistency across individuals to reach statistical significance. While significant effects on a group level are interesting from a fundamental point of view, individual patterns are more relevant, when physiology is applied in human-machine systems. In this respect, the feature extraction and classification algorithms used by BCI researchers offer a promising way of dealing with these limitations.

As shown by Kohlmorgen et al. (2007), driver support may be linked to electroencephalogram (EEG) signals. Given the fact that the driving task is increasingly demanding, due to increased complexity of the road network, increased traffic intensity, and the availability of potentially distracting in-vehicle information systems, such as phones, (e.g., Carsten & Brookhuis, 2005), accurate assessment of user state while driving might be used to benefit driving performance. From driving behaviour literature, it is clear that besides mental workload, other, related psychological constructs might be investigated for use in a support system. At this point there is no consensus about the exact psychological processes underlying driving behaviour. Depending on the theoretical framework, the level of (subjective) risk, workload, or a general feeling of comfort is either maintained or avoided (e.g., risk homeostasis theory, the zero-risk theory, risk allostasis theory, safety margin model (Wilde, 1982; Näätänen & Summala, 1976; Fuller, 2005; Summala, 2005; see also: Lewis-Evans et al., 2011). To make it even more complex, drivers alter the level of workload in practice through behavioral adaptations. For example, in demanding situations with high information density (e.g., complex variable message signs), narrow lanes or a winding road, a driver may reduce speed, which will reduce the reaction time requirements, and thereby avoids high workload levels (Hockey, 2003; Lewis-Evans & Charlton, 2006).

Ultimately, we would like to provide a proof of concept for a passive brain-car interface that changes driving speed in response to visuomotor workload, thereby keeping workload levels within an acceptable range, similar to a human driver. However, in preparation for this, we have first investigated the feasibility of using EEG signals to classify between levels of lane keeping demand in a driving simulator. For this, we applied subject-specific Common Spatial Patterns (CSPs; e.g., Blankertz et al., 2008). The main advantage of using the CSP technique is that it maximizes the difference between two conditions by creating linear combinations of all included electrodes; spatial filters used to produce CSP components. In this way, some electrodes contribute more to the filtered signal(s) than others. These CSP components are determined per participant and therefore, individual differences are accounted for. The most discriminative components are then used to distinguish conditions.

Lane keeping demand was manipulated by changing driving speed, mimicking drivers' natural behaviour. Driving speed was set relative to the participants' comfortable speed, since the effort that is required to keep the car safely on the road may vary between drivers for absolute driving speeds. A relative high driving speed is hypothesized to result in a relative high visuomotor workload. In addition, since the Standard Deviation of the car's Lateral Position (SDLP) reflects workload (e.g., Dijksterhuis et al., 2011), an individually set target SDLP was presented to the participants on the virtual windshield, urging drivers to show less swerving behaviour in the driving lane. A relative low target SDLP is hypothesized to result in a relative high workload level.

## 6.2 Materials and Method

### 6.2.1 Participants

A total of 17 males and 17 females were recruited through social media and poster announcements throughout the University of Groningen and were paid 20 Euros for participation. A large part of the participants were either Dutch or German students at this university. Ages ranged from 21 to 34 years (M=24.0; SD=3.0) and the participants had held their driver's license for 3 to 15 years (M=5.3; SD=2.8). Self-reported total mileage ranged from 3000 to 350,000 km (M=53,000; SD=76,000), while the reported average annual mileage over the past 3 years ranged from 1000 to 50,000 km (M=9000; SD=11,000). None of the participants reported on using prescribed drugs that might affect driving behaviour. The Ethical Committee of the Psychology Department of the University of Groningen has approved the study

### 6.2.2 Simulator and driving environment

The study was conducted using a fixed-base vehicle mock up with functional steering wheel, indicators, and pedals. The simulator runs on ST Software© which is capable of simulating fully interactive traffic. The three computers dedicated to the simulator compute the road environment and traffic which are displayed on three 32-inch plasma screens and provide a total view of the driving environment of 210°. In addition, three rear-view mirrors are projected on the screens. A detailed description of the driving simulator used can be found in Van Winsum & Van Wolffelaar (1993).

For the experiment a two-lane road (each 2.75 m wide) was prepared, without intersections and winding through rural scenery. The road consisted mainly of easy curves (about 80%) with a constant radius of 380 m and ranging in length from 120 to 800 m. The road surface was marked on the edges by a continuous line (0.20 m wide) and in the centre by a discontinuous (dashed) line (0.15 m wide). Outside the edges a soft shoulder was present and there were no objects in the direct vicinity of the road. In the driving direction of the participants, no traffic was present. However, oncoming traffic, travelling between 76 and 84 km/h, was generated with a random interval gap between 1 and 2 s, resulting in 40 passing private vehicles per minute on average. The speed of the participant's vehicle was controlled by the simulator for all rides during the experimental session, except for the initial ride during which the participants drove the simulator car (width: 1.60 m) in automatic gear mode.

### 6.2.3 Design and procedure

Upon arrival at the experimental site of the University of Groningen, a participant was informed in general terms with respect to the experimental design, was requested to sign an informed consent form, and asked to fill in a short questionnaire mainly related to their driving experience. Hereafter, the participant was given some time (ca. 7 min) to practice driving in the simulator, before the sensors were attached. Next, a three minute baseline recording was made while the participant sat in the simulator car chair and an aquatic movie played on the centre screen of the simulator.

After this baseline recording the participant completed 16 short rides. After each ride, the participant was requested to park the vehicle on the side of the road and to provide an answer to two brief questions (on subjective mental effort and estimated driving speed). During the initial ride (140 s) the participant exerted both longitudinal and lateral control over the vehicle and was asked to find and drive at a speed that felt most natural and comfortable in this situation while the speedometer was turned off to prevent rule-based speed setting. The speedometer remained turned off for the entire experiment. The mean speed and standard deviation of the vehicle's lateral position (SDLP) on the road during the last 110 s of the initial ride represented the participant's personal, comfortable driving style. These parameters were saved and used to set driving speed and target SDLP during the 15 remaining rides.

During these 15 rides (130 s each), speed was set relative to the participant's comfortable speed (either -40, -20, 0, +20, or +40 km/h). In addition, while speed was set at the comfortable driving speed, the participant was requested to keep SDLP at either 0, -0.05, or -0.10 m relative to the initial SDLP, which represent a normal, hard, or very hard task. For the other driving speeds, the target SDLP was determined as follows. From a pilot study (n=9), using a similar roadway environment, it was found that SDLP naturally increases as a function of speed. To compensate for this effect and thereby creating five roughly comparable steering challenges across speeds, another 0.03 m per speed level was either added to or subtracted from the target SDLP. For example, when driving 40 km/h slower than the comfortable speed while the target SDLP condition was set at 'very hard', the numerical target SDLP was set 0.10 + 2 × 0.03 = 0.16 m lower than the comfortable SDLP as established during the initial ride. Current values of SDLP were derived from a 15 s moving window which was updated every second and these were projected onto the bottom of the windshield of the simulator while driving, adjacent to the target SDLP. In this way a driver could monitor real SDLP and compare it to the target. Accounting for the time window and for the time the simulator needed to get to the required speed, only the last 110 s of each ride was used in subsequent analyses. To be clear, the data used for this analysis were the raw, not averaged, vehicle parameters. In total, the experimental manipulations resulted in a within-subject design consisting of two repeated measures factors with several levels: speed (5) and target SDLP (3). The participants were exposed to these driving conditions according to a randomized schedule.

After finishing the last ride, the baseline measurement was repeated once more before all physiological sensors were removed. Finally, the participants were debriefed and were paid upon leaving.

### 6.2.4 Dealing with collisions

Occasionally, the participants were challenged to the point that a collision with oncoming traffic could not be avoided. In total, six participants were involved in 10 collisions which is

1.8% of all experimental rides. Eight of these collisions occurred in a +40 km/h speed condition. When a collision occurred, that particular ride was repeated. Data acquired during the crash rides were not used for further analyses.

### 6.2.5 Data acquisition

#### 6.2.5.1 Vehicle parameters

Driving speed and lateral position (LP) were sampled at 10 Hz. LP is defined as the difference in meters between the centre of the participant's car and the middle of the (right hand) driving lane. Positive LP values correspond to deviations towards the right hand shoulder and negative values correspond to deviations towards the left hand shoulder. The sampled LP values were processed while driving and used to calculate mean LP and SDLP for each of the 16 rides. In addition, LP values were used to feed current values of SDLP back to the participant which were calculated by using moving, overlapping time windows (see 6.2.3 Design and Procedure for more details), representing an indication of the participants lane keeping performance.

#### 6.2.5.2 Subjective ratings

After each ride, a rating on the one-dimensional Rating Scale Mental Effort (RSME) was requested (Zijlstra, 1993). The RSME ranges from 0 to 150 and several effort indications are visible alongside the scale which may guide the participant in marking the scale. Indications start with 'absolutely no effort' (RSME score of 2) and end with 'extreme effort' (RSME score of 112). The participants, who did not receive speed information from the speedometer, were also asked to write down an estimate of the driving speed they just experienced.

#### 6.2.5.3 Physiological measures

Physiological signals were sampled at 250 Hz. Firstly, the electrocardiogram (ECG) was registered using three Ag-AgCl electrodes, which were placed on the sternum (the ground electrode) and on the right and left side between the lower ribs. However, given the emphasis on brain activity in this paper, the ECG results are not reported here. Secondly, the electro-oculogram (EOG) was measured by Ag-AgCL electrodes attached next to the lateral canthus of each eye and above and below either the right or left eye. The EEG was measured using an electro-cap with 64 tin electrodes (at the following sites: FP1, FP2, Afz, F7, F5, Fz, F4, F8, T7, C5, C3, Cz, C4, T8, P7, P3, Pz, P4, P8, O1, and O2.)[1] The amplifier was a REFA 8–72 (Twente Medical Systems International, Enschede, The Netherlands). Portilab 2 software was used to record all physiological signals. The ground electrode used for the ECG recording also served as the participant's ground for the EEG recording. EEG and EOG signals were amplified with a 1 s time constant (0.016 Hz high-pass). All EEG channels were referenced against the average activity of all channels during the registrations. In addition, a reference electrode was attached to each mastoid. Impedances were kept below 10 kΩ for all electrodes.

### 6.2.6 EEG data processing

Starting from the raw EEG signals, the sampled EEG and EOG data were first high-pass filtered (cut-off = 0.3 Hz, at 12 dB/Oct Butterworth filter) before the EEG data segments of the 15 experimental conditions (110 s each) were corrected for eye movements and blinks, using Brain Vision Analyzer (Gratton et al., 1983). The corrected data segments were then exported into binary files. No data epochs were removed before further processing.

The exported data files were processed using MATLAB R2010a (The MathWorks, Inc., USA, www.mathworks.com). After importing two data sets (two rides or conditions) of a particular participant, the EEG was band-passed filtered in the frequency domain (FFT filter) of interest, using an edge frequency of 1 Hz below and above the lower and upper frequency band limit respectively. The imported data (110 s for each condition) were then segmented into one second epochs and baselined relative to each mean activity. The first and last 10% of the epochs were omitted, leaving the 88 middle, non-overlapping, epochs per condition in the cross-validation design. This entailed a repeated (50 times) random portioning of two data classes (a condition pair) into a set of 66 training epochs (75%) and a set of 22 test epochs for each data class. The training sets were used to train the participant-specific classifier that was subsequently used to classify the testing epochs of each data class. This iteration process was carried out for each included participant, frequency band, EEG cap configuration, and data pair. The accuracies reported in the result section reflect the average accuracies across all 50 iterations and all included participants.

To improve discriminatory power of the data classifier, the contrast between two data classes was optimized by using the CSP technique. This technique determines CSP filters in such a way that they maximize the variances of spatially filtered signals for one training set while

[1] The included electrodes were based on the International 10-20 system. However, C5 was added based on literature on the 'engagement index' (Pope et al., 1995). Using F3 would have been in accordance with the 10-20 system. However it is unlikely that using F5 instead had any consequences for the main results of the current study.

minimizing them for the other (Blankertz, 2008). A CSP filter is a coefficient vector by which the original channels can be transformed. This results in a new spatially filtered channel (a CSP component) which is a linear combination of all original channels, and the total number of filters and therefore, the number of components, is equal to the number of original channels. The matrix of CSP filters is determined by solving a generalized eigen-value problem. The filter corresponding to the largest eigen-value yields a high variance signal in one condition, while producing a low variance signal in the other; and vice versa for the filters corresponding to the smallest eigen-value. The CSP filters are therefore ranked according to these eigen-values and the first and last filters in this sorted W matrix are usually used for further classification. To be more specific, in the current study, the two, four, or six filters (always an equal number from each side of the sorted W matrix) that resulted in the largest difference in variance between two training sets was used. Next, the total variance per training epoch and per CSP component was calculated and their logarithms were taken before entered into Fisher's linear discriminant analysis. This analysis again transforms the data by determining the linear weights of the discriminant function that combines data points of the two training sets in such a way that maximizes the distance between them. Finally, the CSP filters and classifier weights were used to classify the remaining testing epochs of the two conditions.

A wide range of EEG frequency bands were explored to investigate where useful discriminatory information might be present. Four frequency search strategies were deployed. The first frequency search strategy was characterized by both an increasing high pass cut-off point (increasing 1 Hz for each iteration) and an increasing frequency bandwidth (1.5 times the low frequency band limit). At the first iteration, frequencies between 3 and 4.5 Hz were passed. At the last iteration, frequencies between 72 and 108 Hz were passed. The second strategy entailed exploring all frequencies between 3 and 70 Hz using a fixed bandwidth of 1 Hz. For the third strategy, bandwidth was set to 4 Hz and iterations ran from 4 to 72 Hz. Lastly, data was filtered in broad bands to classify between conditions; 8-30 Hz, 32-54 Hz, 56-78 Hz, and 80-102 Hz.

In addition, several EEG cap configurations were explored. To start with, all 21 EEG channels were included. To explore whether classification accuracy may differ between scalp regions, several subsets were defined and tested. Firstly, a peripheral set was defined, consisting of 14 electrodes, (FP1, FP2, Afz, F7, F5, F4, F8, T7, C5, T8, P7, P8, O1, and O2). Secondly, a frontal set consisting of 7 electrodes (FP1, FP2, F7, F5, Fz, F4, and F8), which are associated with executive functions that are important in driving. Thirdly, a posterior set consisting of 7 electrodes (P7, P3, Pz, P4, P8, O1, and O2), containing electrodes associated to visuomotor processing. Lastly, the electrode set identified by Prinzell et al. (2001; P3, Pz, P4, Cz), which has often been used in adaptive automation research to get the 'engagement index' (defined as the ratio; beta/(alpha + theta)).

Lastly, five condition pairs were selected from a total of 105 possible combinations (15!/2!(15-2)!). An experimental condition can be defined in terms of its driving speed level and target SDLP difficulty level. To improve comparability one factor was kept constant for each condition pair. In this way, four speed differences for the normal target level were classified: -40 vs. +40 km/h, -20 vs. +20 km/h, -20 vs. 0 km/h, and 0 vs. +20 km/h. The normal target level was chosen since this target resembles the individuals' natural driving behaviour. Focusing on classifying between speed differences in this way was done because of the envisioned application. A brain-based adaptive cruise control would change speeds and therefore, the effect of speed interventions has to be assessed. In addition, as it turned out, the very hard target conditions required more subjective effort compared to the normal target level, and therefore, these two conditions were compared in the 0 km/h relative speed condition.

Due to data anomalies such as missing channels, eight participants were excluded from the offline classification phase of this study. Despite a smaller participant pool, the number of condition pair comparisons is very large: 161 frequency bands × 5 condition pairs × 5 EEG cap configurations × 26 participants × 3 numbers of components = 313,950. Given these large numbers, only a selection of aggregated classification accuracy values can be reported (Figure 6.2 and Figure 6.3) next to examples of scalp topographies of CSP components (see Figure 6.4 for an impression) reflecting how the information sources project to the scalp (retrieved from the inverse of W; see Blankerz et al., 2008).

## 6.3 Results

### 6.3.1 Vehicle parameters and subjective ratings

Subjective ratings and vehicle parameters are shown in Figure 6.1 and their test outcomes are listed in Table 6.1. To begin with, the participants' preferred speed during the initial ride ranged between 62 and 120 km/h, averaging at 90 km/h (see the black dot in Figure 6.1A). This is slightly faster than estimated for this ride (M =74 km/h; Figure 6.1B). This pattern of underestimating driving speed is present for all speed levels (Pearson's product moment correlation = 0.99 over all conditions).

The dimensions of the vehicle and driving lane allowed for 0.58 m of swerving margin on both sides of the vehicle. As can be seen in Figure 6.1C, the participants stayed well within their driving lane on average and positioned the vehicle slightly towards the right hand shoulder (0.07 m on average). As can be read in Table 6.1, there was a significant effect of speed on LP. The participants' mean position on the road curves toward the right-hand shoulder, both when driving slower and faster than the preferred speed (polynomial contrasts showed a quadratic trend; ($F(1,33) = 15.35$, $p < 0.001$, $\eta_p^2 = 0.317$). Next, as speed increased, so did the participants' mean SDLP (see Figure 6.1D), representing swerving behaviour, from 0.18 m during the slowest speed to 0.30 m during the fastest speed. This is mainly a linear increase ($F(1,33) = 182.81$, $p < 0.001$, $\eta_p^2 = 0.847$), although SDLP increases slightly more rapidly towards the higher speeds (quadratic trend; $F(1,33) = 24.33$, $p < 0.001$, $\eta_p^2 = 0.424$). Note that the factor; target SDLP, indicating the difficulty of keeping current SDLP values under the target

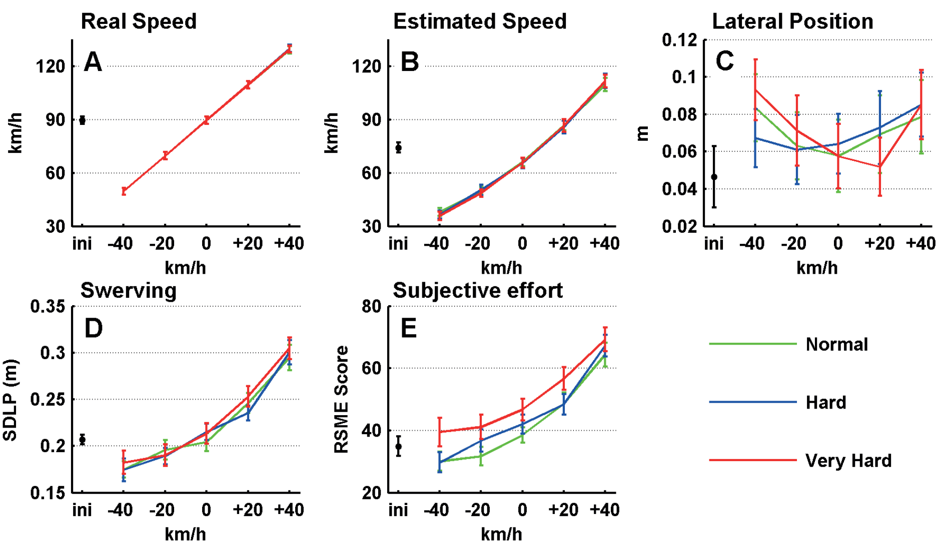



**Figure 6.1.** Vehicle parameters and subjective ratings as a function of set driving speed condition. (A) Real driving speed. (B) Estimated driving speed. (C) Lateral Position (LP). (D) Standard Deviation of the Lateral Position (SDLP). (E) Rating Scale Mental Effort (RSME). On the x-axes, values for the initial ride (black dots) are shown in addition to five driving speeds that were set, relative to the individual's preferred driving speed established during the initial ride. Error bars represent the standard error. LP values represent the middle of the car (car width = 1.60 m) in relation to the middle of the right (driving) lane (width = 2.75 m). Normal, hard, and very hard indicate the difficulty of keeping current SDLP values under the target SDLP: see Section 6.2.3 for details. Positive LP values indicate a position to the right hand of the lane mid. Maximum score for mental effort is 150. n=34.

SDLP while driving, had no effect on the actual SDLP. In addition, interactions between speed and target SDLP are not present in the data.

Figure 6.1E shows that the mental effort ratings increased from between ‘a little effort’ and ‘some effort’ (a mean RSME score of 33) for the slowest speeds to between ‘rather much effort’ and ‘considerable effort’ (a mean RSME score of 34) for the fastest speeds (linear trend; $F(1,33) = 88.48$, $p < 0.001$, $\eta_p^2 = 0.728$). Also, similar to SDLP, this increase is stronger towards the faster speeds (quadratic trend; $F(1,33) = 86.04$, $p < 0.001$, $\eta_p^2 = 0.327$). In addition, even though target SDLP did not have an effect on vehicle parameters, there was a main effect on mental effort ratings. Bonferroni corrected pairwise comparisons revealed that the ‘very hard’ level was perceived as more difficult than the other two, while ‘normal’ and ‘hard’ did not show a difference.

**Table 6.1.** Multivariate test results for vehicle parameters and subjective effort ratings (Figure 6.1). LP = Lateral Position, SDLP = Standard Deviation Lateral Position. RSME = Rating Scale Mental Effort. Significant effects (p < 0.05) are shown in bold. Speed effect relates to speed condition, Target to SDLP target.

| **Vehicle parameters and subjective ratings** | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | **LP** | | | **SDLP** | | | **RSME score** | | |
| *Effect* | F(df1,df2) | p | $\eta_p^2$ | F(df1,df2) | p | $\eta_p^2$ | F(df1,df2) | p | $\eta_p^2$ |
| **Speed (S)** | **4.46 (4,30)** | **0.006** | **0.373** | **45.40(4,30)** | **<0.001** | **0.858** | **21.10(4,30)** | **<0.001** | **0.748** |
| **Target (T)** | 0.26(2,32) | 0.974 | 0.002 | 1.32(2,32) | 0.283 | 0.076 | **8.49 (2,32)** | **0.001** | **0.347** |
| **S × T** | 0.77(8,26) | 0.633 | 0.191 | 1.22(8,26) | 0.324 | 0.274 | 1.25(8,26) | 0.309 | 0.278 |

### 6.3.2 Classification results

#### 6.3.2.1 Averages classification accuracies

In Figure 6.2, the classification accuracies for several condition pairs are shown. Figure 6.2 (and Figure 6.3) only shows the average classification accuracies for two data pairs (-20 km/h vs. +20 km/h and normal performance target vs. very hard performance target). Although more extreme driving speed conditions could have been shown (e.g., 40 km/h vs. +40 km/h), we feel that more similar speed conditions better reflect real driving circumstances and are therefore more relevant. Also, accuracy levels across condition pairs tended to be similar, and therefore the number of shown condition pairs was limited.

The graphs in Figure 6.2 reveal several general trends. Firstly, accuracy tends to increase as frequency increases. This can be seen across electrode sets and condition pairs with accuracies reaching levels of 95% on average over all participants when a relative high number of electrodes is included (21 and 14). This increase is most pronounced in the frequencies from 5 to 20 Hz, after which it continues to rise more gradually indicating a ceiling effect (see for example Figure 6.2A). This ceiling is about 5-10% lower for the middle column of subplots in Figure 6.2 (displaying the 4 Hz search strategy). Secondly, a broader frequency band tends to yield higher accuracies, which is most apparent when comparing the middle column (Figure 6.2B and Figure 6.2E; 4 Hz frequency bands) to the right column (Figure 6.2C and Figure 6.3F; 22 Hz frequency bands). For example, when including all electrodes, the 4 Hz frequency bands in the 8-32 Hz range in Figure 6.2B range produced about 15% less accuracy when compared to the first broad band (8-30 Hz) in Figure 622C. Thirdly, there are distinct differences in accuracies as a result of using different channel sets. For example, the larger electrode sets (21 and 14 electrodes) yielded very comparable high accuracies, while the smallest (4 electrodes) consistently resulted in lower classification accuracies (about 15-25% less, depending on frequency band). Such differences can be understood in part from the fact that more channels provide a richer, higher-dimension database for the CSP technique to extract useful discriminatory power. Note however, that the seven frontal electrodes outperformed the seven posterior electrodes by about 5-15%, again depending on frequency band. The shape of the frontal curve in all subfigures (the red lines) reflect the upper two curves (all electrodes and 14 peripheral electrodes), while the posterior curves resemble the bottom El curves. Finally, when focusing on the somewhat lower EEG frequency of Figure 6.2A and Figure 6.2D ranges (e.g., 10 to 21 Hz), which are more likely to reflect neuronal

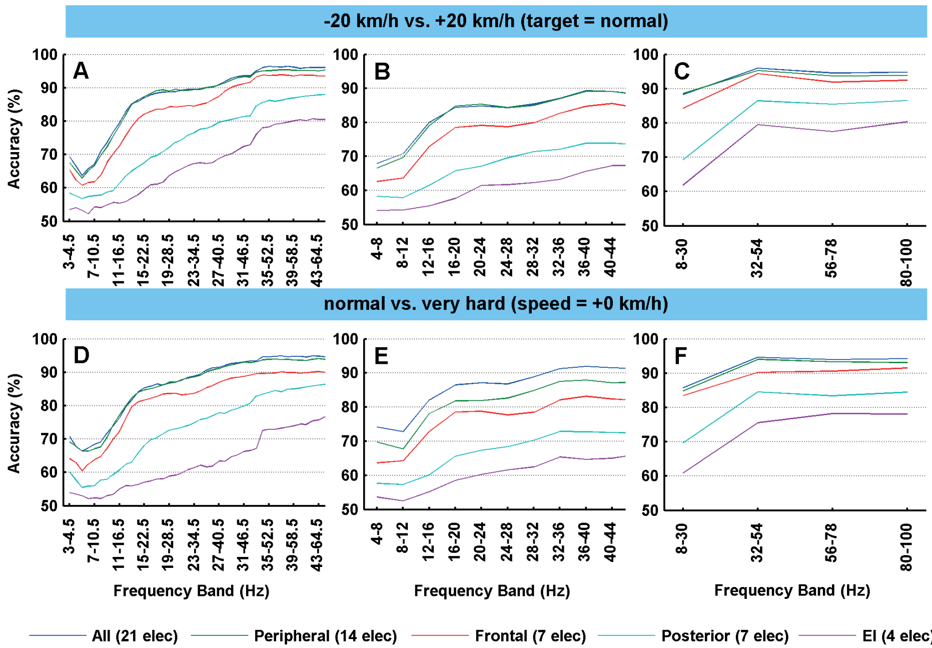



**Figure 6.2.** Average classification accuracies of the Fisher's linear discriminant analyses after spatial filtering for several condition pairs. (A-F) The accuracy values represent the average subject-specific classification accuracy over all participants that resulted from the cross-validation scheme. Classifications were based on applying the two most contrasting CSP components to the EEG channels. N=26. For each row of subfigures, a different EEG cap configuration was used. For the left column (A,D), the frequency bandwidth is 1.5 times the start frequency (step size 1 Hz), starting at 3-4.5 Hz and ending at 43-64.5 Hz. For the middle column (B,E), 4 Hz bands were used and a step size of 4. For the right column (C,F), a broad band frequency search (22 Hz) was deployed. All electrodes: FP1, FP2, Afz, F7, F5, Fz, F4, F8, T7, C5, C3, Cz, C4, T8, P7, P3, Pz, P4, P8, O1, O2. Peripheral set: FP1, FP2, Afz, F7, F5, F4, F8, T7, C5, T8, P7, P8, O1, O2. Frontal set: FP1, FP2, F7, F5, Fz, F4, F8. Posterior set: P7, P3, Pz, P4, P8, O1, O2. Engagement index (EI) set: P3, Pz, P4, Cz.

activity, the mean accuracy in that range over both subfigures is 80% for the larger two electrode sets. The frontal set led to a classification of 76% on average, while the posterior and the engagement index set resulted in 62% en 55% respectively.

#### 6.3.2.2 Cumulative classification accuracies

In Figure 6.3, cumulative classification accuracies are shown for a selection of classification results. This figure indicates the consistency of classification accuracies across all 26 included participants. For instance, Figure 6.3A shows that in the high frequency range (e.g., 43-64.5 Hz), test data from 10 participants were accurately classified 99% of the time or better. Figure 6.4 confirms that higher frequencies usually yield better accuracies as the top frequencies in all subfigures display more red/yellow than the bottom frequencies. The green/yellow colours indicate that about half to two third of the participants were above the classification threshold indicated on the x-axes. When viewing these colours in Figure 6.3 through the eyelashes, it can be seen that, especially for the larger electrode set (Figure 3A,E,I and Figure 6.3C,G,K), data from a substantial number of participants still yielded 85%+ accuracy in the lower (alpha and beta) frequency ranges (e.g., 10-20/30 Hz). For instance, the classifier could accurately classify (85% or better) between -20 and +20 km/h in the 16-20 Hz frequency range for 16 out of 26 participants (Figure 6.3G). For the smaller, frontal electrode set (Figure 6.3B,F,J and Figure 6.3D,H,L) the number of participants yielding highly accurate classifications is somewhat less in the lower frequency range; as indicated by the larger presence of blue colours.

#### 6.3.2.3 Example common spatial pattern analysis

Figure 6.4 displays several CSP filter-topography pairs which are meant to illustrate the diversity of CSP scalp topographies. A common topography across participants, reflecting how the neurological sources project to the scalp, was not identified. However, we selected these topographies based on their resulting classification accuracies and/or the fact that the frequencies are within the normal EEG range. To start with, Figure 6.4A shows that for participant 13, the perfect classification accuracy in the broad 72-108 frequency range originates mainly from the frontal electrodes (Fp1 and Fp2) which were highly specific for the -20 km/h driving condition, and from C5 which was highly specific for the +20 km/h driving condition. This is illustrative for the general finding that the frontal electrodes were often the main contributors to very high classification accuracies. The other subfigures show topographies linked to frequencies below 30 Hz. In Figure 6.4B, topographies are shown that resulted in an unusually accurate classification for this relative low frequency band (98% in the

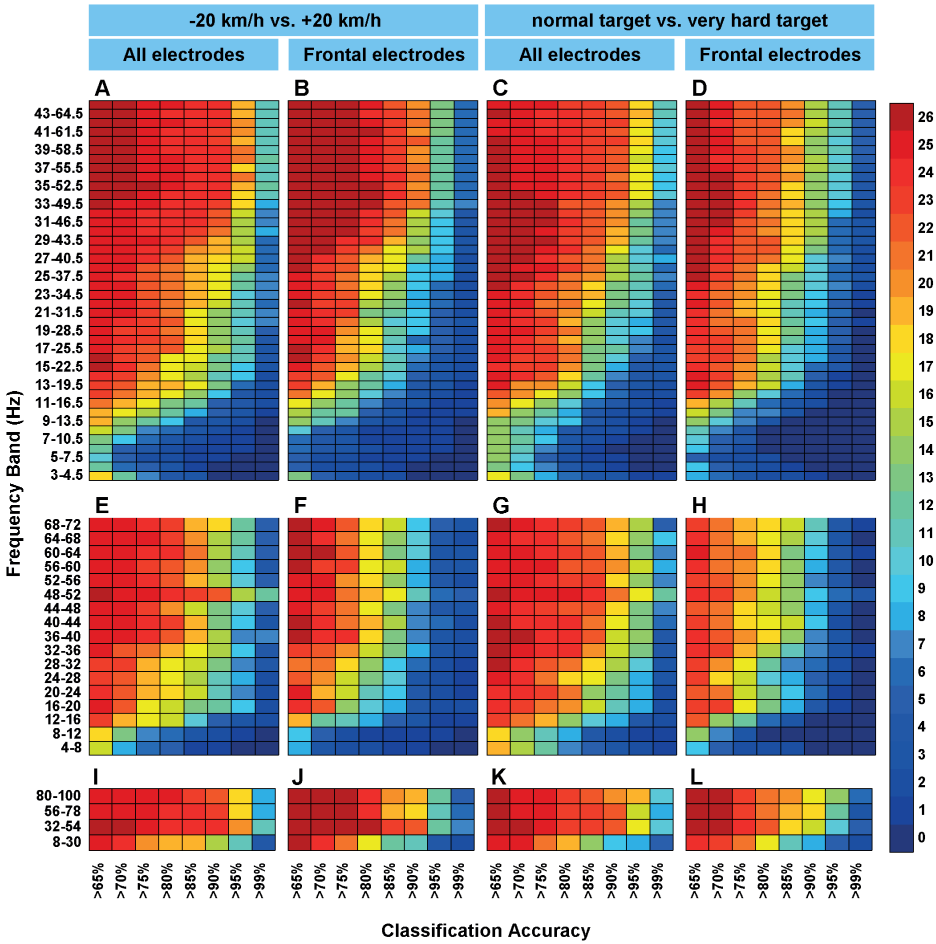



**Figure 6.3.** The cumulative frequencies of classification accuracies. (A-L) Colours represent the number of participants for whom a particular accuracy was found or better (max = 26 participants) in the accuracy category displayed on the x-axes. Subfigure columns are arranged by EEG cap configuration (all electrodes or the frontal set) and by classified condition pair. All electrodes: FP1, FP2, Afz, F7, F5, Fz, F4, F8, T7, C5, C3, Cz, C4, T8, P7, P3, Pz, P4, P8, O1, and O2. Frontal electrodes: FP1, FP2, F7, F5, Fz, F4, and F8. '-20 km/h vs. + 20 km/h' indicate set driving speeds relative to the participants' preferred speed as determined during the initial ride (at normal target level). 'Normal target vs. hard target' indicate performance target difficulty (at relative speed = 0 km/h). For each row of subfigures, a different frequency search strategy was used. (A-D) For the top row of subfigures, the frequency bandwidth is 1.5 times the begin frequency (step size 1 Hz), starting at 3-4.5 Hz and ending at 44-66 Hz. (E-H) For the middle row of subfigures, 4 Hz bands were used and a step size of 4. (I-L) For the bottom row, a broad band frequency (22 Hz) search was deployed.

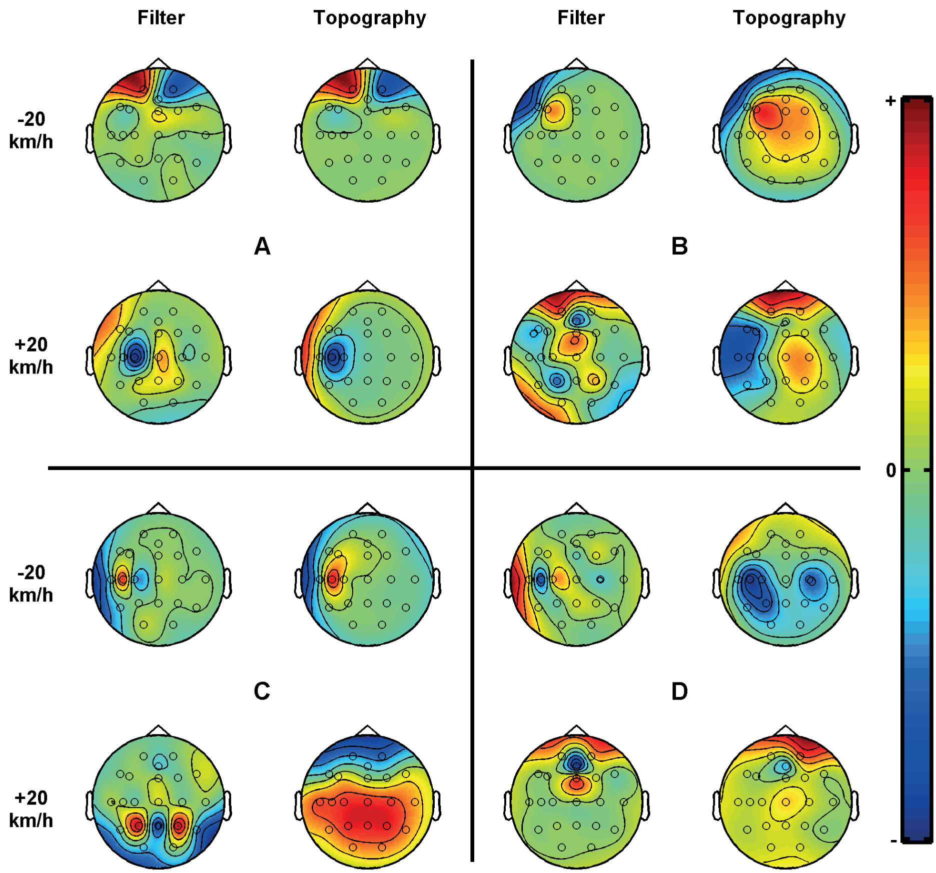



**Figure 6.4.** Examples of CSP analyses. (A-D) The scalp topography of the components illustrate how the physiological sources project to the scalp. The components are determined such that projected signals are optimally discriminative. The filters and topographies correspond to the first and last vector of the sorted W matrix and its inverse respectively (see Section 6.2.6 for more details). Absolute colouring is arbitrary, however dense red or blue areas show where the greatest differences in the projected signals' amplitudes were found, between the -20 km/h and the +20 km/h set driving speed (at normal target level). These driving speeds were set relative to the participants' preferred speed as determined during the initial ride. Included electrodes: FP1, FP2, Afz, F7, F5, Fz, F4, F8, T7, C5, C3, Cz, C4, T8, P7, P3, Pz, P4, P8, O1, and O2. A: Subject = 13, frequency band = 72 -108 Hz, classification accuracy = 100%. B: Subject = 27, frequency band = 8-12 Hz, classification accuracy = 98%. C: Subject = 21, frequency band = 24-28 Hz, classification accuracy = 82%. D: Subject = 25, frequency band = 8-30 Hz, classification accuracy = 74%. Please note that in the CSP literature, a scalp topography of a component is usually referred to as a spatial pattern.

8-12 Hz, alpha, frequency band). In this case, the topographies are more distributed over the scalp, although the left temporal and frontal regions were important physiological sources for discriminating between the two data classes. The scalp topography for the +20 km/h condition in Figure 6.4C shows a central-parietal distribution, illustrating that the EI electrodes: P3, Pz, and P7 contributed to the 82% classification accuracy in the high beta range of participant 21. The -20 km/h topography suggests that C5 was by far the most distinctive electrode when maximizing the variance of the projected signals in this data class while minimizing it for the other. Finally, Figure 6.4D shows the CSP resulting in 74% classification accuracy for subject 25 in the broad 8-30 (alpha plus beta) Hz frequency band. These topographies suggest that discriminative power was distributed over the posterior electrodes in the -20 km/h condition and more evenly distributed over the scalp in the +20 km/h condition.

## 6.4 Discussion

The aim of the study was to investigate the feasibility of using EEG for monitoring the level of visuomotor workload in a driving environment, which can potentially be used by an user adaptive driver support system. To manipulate workload, we exposed drivers to five levels of driving speed that were set relative to their preferred driving speed. In addition, since increasing steering effort normally decreases swerving behaviour within the driving lane given a particular speed, participants were presented with three explicit swerving performance targets represented as the standard deviation of the lateral position of the car with respect to the driving lane. To distinguish between workload levels, subject-specific CSP and linear discriminant analysis based classification models were used.

To begin with, subjective mental effort data show that driving at a higher speed is indeed experienced as requiring more effort. Furthermore, estimated driving speed was slightly lower than the real driving speed. Previous research has shown that driving speed in a simulator, when driving on straight roads or easy curves, tends to be higher than it would be on real roads (e.g., Bella, 2008). This effect could be caused by a difference in speed perception between the real world and (fixed-base) driving simulators due to the absence of several speed cues, such as car movements and stereoscopic depth perception. During the current study these factors may also have contributed to misjudging driving speed, especially since the speedometer was hidden from view at all times. The standard deviation of the lateral position (SDLP), indicating lane keeping performance, increased as function of speed, which is normal (e.g., Peng et al., 2013). However, the performance target (target SDLP) did not have an effect on vehicle parameters, suggesting that this manipulation failed since a decrease of SDLP was expected if participants were exposed to more difficult target SDLPs. However, participants did rate the 'very hard' target SDLP condition as the most difficult, perhaps demonstrating that participants were trying hard but could not manage. Also, EEG data from the very hard SDLP condition could be accurately discriminated from data acquired during the normal SDLP condition which is another indication that participants did not simply ignore the instructions. Since other task manipulations aimed at increasing steering difficulty, such as decreasing lane width, have proven to affect SDLP (e.g., Dijksterhuis et al., 2011), the absence of an effect on SDLP may be explained by this particular manipulation. In contrast to the automatic nature of the steering task during normal driving participants had to actively engage themselves in transferring numerical information about their lane keeping behaviour, as presented on their windshield, to steering wheel movements.

EEG activities during the experimental conditions were classified, yielding several interesting results. Firstly, applying CSP to a variety of frequencies and frequency band widths revealed that, overall, broader bands and higher frequencies result in higher classification accuracies. This could be taken to suggest that neuronal gamma synchronization correlated with the task manipulations in which case these results are in line with other research suggesting that activity in the gamma frequencies reflects sensory-motor coordination (Schoffelen et al., 2005, see also Fries et al., 2007). However, this conclusion should be drawn with caution since muscle activity as represented in the EMG has power in the same frequencies, which is picked up by EEG electrodes as well (Whitham et al., 2007; Muthukumaraswamy & Singh, 2013). This view of muscular activities contributing to high classification accuracies in the gamma band is confirmed by graphs showing the projections of the CSP components. Figure 6.4a demonstrates just one case where the perfect classification for high frequencies can mostly be traced to the EEG electrodes close to the eyes. However, a relative high contribution of the peripheral electrodes for extremely high classification accuracies is an emerging pattern. Moreover, when performing a semi-real task, such as driving in a simulator, EMG activity can be expected to be more dominantly present compared to more controlled laboratory tasks, making classifications based on neuronal gamma activities less likely.

High accuracies were also found for a substantial number of participants in the lower frequency ranges, as shown in Figure 6.2 and Figure 6.3, and these are unlikely to be confounded by EMG activity. As shown by Whitham et al. (2007), who recorded EEG during paralysis by neuromuscular blockade, EMG activity is largely absent from frequencies below 20 Hz. Therefore, we suggest that classifications in the lower frequency ranges were likely determined by underlying neuronal activity.

CSP component topographies showed no readily discernible degree of consistency across participants, as illustrated in Figure 6.4. This indicates that the effects of changes in psychological construct such as mental workload on electrical activities on the scalp is very subject dependent, which confirms that individually tuned classification approaches are required for accurate classifications. In case of high frequencies this implies that the, perhaps subconsciously produced muscular activities, show large inter-individual variations. In case of the lower frequencies, it is likely that also on a neurological level, there are large variations. Finding consistent topographies would have been promising for future applications. For example, it could lead to a theory-driven pre-selection of scalp locations, thereby excluding possible irrelevant information from the classification model. Yet, it may be expected to find a large inter-individual variability when classifying rather abstract mental states compared to, for example, classifying the difference between left and right hand motor imagery for which the neuroanatomical base is much clearer.

A limitation of the current study is that the experimental conditions (rides) could not be randomized within each participant. E.g., changing speed conditions every couple of seconds would have resulted in a highly unnatural driving experience. The drawback of the used approach is that there was an average of about 15 minutes between one condition and the other within each condition pair that was used for the classifications. Since neighbouring epochs can be similar to each other, a difference in time may have led to an inflation of the classification accuracies. For future research, it is advised to repeat conditions within subjects to assess the potential effects of time dependencies. For example, by training the classifier on one condition pair and validating it on the other, identical condition pair. While it is important to realize that time dependencies cannot be ruled out, it should also be noted that it probably did not affect other effects, such as the accuracy difference between the parietal and frontal electrode set or the difference between low and high EEG frequencies.

Overall, these findings imply that the subject-specific CSP approach provides very good discriminatory power between visuomotor workload conditions over a large range of frequency bands. With respect to the high (gamma) frequency ranges it is important to realize that major contributions from muscular activities cannot be ruled out. Moreover, this will probably be true for most passive BCI applications as real life tasks, such as driving a car, usually require a lot of motor activity. A workload classification strategy based on EMG activity would therefore be worthwhile investigating in future research, which requires a relative low number of electrodes. However, high classification accuracies were also found for the lower EEG frequencies, implying a large contribution of neurological activities. These high accuracies are promising for future applications, however, several issues need to be addressed before a system is working from the user's point of view. Some of these issues will be further discussed below.

Even if classification accuracies of up to 80% may be considered quite high for one-second epochs, it raises the issue of applicability; especially when performing a safety-critical task, this seems insufficient. However, depending on the temporal responsiveness requirements of an application, these accuracy levels might suffice. For example, using longer data epochs can be expected to result in more accurate classifications, since more information is available to the classifier (e.g., Brouwer et al., 2012). Although not further reported in the result section, increasing the epoch length from one to two seconds was found to increase accuracies with about 3% for the lower frequency ranges. Another option would be to combine several successive small data epochs. As an illustration, assuming that successive classifications are independent and applying a simple binomial chance distribution, then combining five successive epochs, each having a 80% chance of accurately being classified, would lead to a 94% accuracy when using a majority vote (i.e., three or more epochs are classified correctly). This would decrease the negative effects of small periods of noisy data which may be expected in real life tasks and which should improve a system's behaviour from the users point of view.

Another important issue that needs to be solved before reliable applications can be build are the so-called non-stationarities in EEG signals, which refer to shifts in EEG signals between the initial calibration session during which a model is trained and online application. Non-stationarities negatively impact the transfer of classification accuracies between calibration and application of a model (e.g., Shenoy et al., 2006). One solution to this issue could be to update the classification model from time to time by adding additional calibration periods when the task at hand allows for it. Another solution are adaptive classifiers, which use data that are acquired while the user is interacting with the system in real-time (Shenoy et al., 2006). The drawback of using an adaptive classifier is that it requires immediate labelling of new, incoming data while the user is engaged in task performance. In some active BCI systems, for example, when controlling a game, it is plausible that the required information is available. In case of a passive BCI system however, this is most likely not the case. Again, using longer periods of time may offer a solution to this problem. For example, assuming that mental workload does not vary every second, all EEG data measured over a somewhat longer period reflect one particular level of workload. If the classifier therefore classifies most epochs as data class A, then all epochs in that period could be labelled as such and subsequently used to update the classification model. Finding acceptable and robust methods of updating the classification model is likely to be a necessary development before (passive) BCI systems can be applied to task situations.

For the viability of future applications it is also important that the binary approach of discriminating between two data classes is expanded to the multiclass situation. For instance, workload levels during task performance may be either too high, too low or within an acceptable range. In an adaptive system, where support may be changed, activated, or deactivated based on workload classifications it is therefore of equal importance that the conditions for no change are defined. Thus, in terms a passive BCI application, a homeostatic system aimed at keeping workload at or around optimal levels, must also 'know' when not to initiate changes. One way to accomplish multi-class analyses is to combine several pairwise classifications through voting procedures (Friedman, 1996; see also Dornhege et al., 2004; Grosse-Wentrup & Buss, 2008).

In conclusion, depending on temporal responsiveness requirements, a system's designer may have the option to either focus on high EEG frequencies and accept that muscular activities likely contribute to classification accuracies, or to focus on lower EEG frequencies that mainly reflect neurological activities but accept slightly lower accuracies. Although it is clear that the very high classification accuracies found in this offline study by themselves do not guarantee a well-functioning online system, it is a promising start in realizing a CSP based passive BCI system that can reliably be used to monitor visuomotor load in real-time.

## Acknowledgments

Firstly, the authors would like to thank graduate students Lucas van Assen and Camilla de Jong for their contributions in recruiting participants, carrying out the experiment, and data processing. Secondly, the technical support given by Mark Span is greatly appreciated. Thirdly, the reviewers are thanked for their helpful suggestions during the publication process. Finally, this research was supported by the European Commission Project No. 215893 REFLECT, Responsive Flexible Collaborating Ambient, within the 7th framework programme. The REFLECT project investigated ways of realizing pervasive-adaptive environments.

# Chapter

# 7

# A brain and performance based adaptive cruise control

**Abstract**

In contrast to an active Brain-Computer Interface (BCI), where users are engaged in volitional thought control of a device, several BCI researchers have proposed to advance human-machine interaction by adapting computer actions to the automatically inferred mental state of its user: passive BCI. Future cars could use passive BCIs to monitor the driver's workload and intervene to avoid extreme levels thereof for example. To investigate the plausibility of this statement, 18 drivers completed several simulator rides on rural roads, which varied in curve length to manipulate lane-keeping demand, while a BCI was monitoring their visuomotor workload. The BCIs were either trained on low or high frequency filtered EEG signals acquired during a low, comfortable, and high load calibration ride. Common Spatial Pattern (CSP) filters and the linear discriminant functions were determined for each pair of workload levels and subsequently used during the application phase. During the application phase, multiple pairwise comparisons, a voting procedure, and an exponential weighted moving average were used to establish the workload level of each second of new data. In addition, vehicle parameters were monitored and in case indications of risky behaviour were not detected, the BCI loop was allowed to determine driving speed to keep visuomotor workload around the comfortable level. Results indicate high classification accuracy for the calibration data, especially for the high frequency models. Although a later discovered technical problem somewhat limits the scope of the conclusions that can be drawn from the application phase of this study, it can be reported that the system's behaviour showed large variations from condition to condition. It is concluded that research into improving the transfer of classification accuracy between model training and model implementation is necessary before passive BCI systems can be reliably used.

## 7.1 Introduction

Interacting with technology is an integral, substantial part of our daily lives. We usually interact with a device by manipulating its interfaces, such as moving pedals, levers, or by issuing voice commands. However, computing power has advanced to the point that huge amounts of additional data from other information sources may be processed and analysed in real-time. For example, biometrical data may be acquired while we are engaged in human-machine interaction (HMI) and can be used to automatically infer our emotions, cognitions, and other mental states. User state information could then be used by the HMI system to change its behaviour to accommodate us (Brusilovsky, 2001; Duric et al., 2002; Hettinger, 2003; Feigh et al., 2012). For example, a computer may offer extra assistance with performing a task when biometrical data show that a user is overloaded. One way of inferring relevant user states is through the feature extraction and classifying power used by Brain-Computer Interface (BCI) researchers (passive BCI; Kohlmorgen et al., 2007; Curtrell & Tan, 2008; Zander et al., 2010; Zander & Kothe, 2011). For example, Dijksterhuis et al, (2013) showed that it is possible to accurately discriminate between high and low visuomotor workload, elicited by speed changes on rural road conditions in a driving simulator.

Inferring user state in real time has been a major topic of investigation in another subfield of HMI as well. In the human factors and ergonomics literature, the potential detrimental effects of extreme workload levels have been well documented. While humans are often able to deal with short bursts of task demand, when exposed to longer periods of high workload it can lead to performance degradation, such as neglecting secondary tasks and ultimately to a performance breakdown of the primary task (e.g., Hockey, 2003, 2007). Therefore, developing task environments aimed at keeping the human workload within acceptable range by dynamically allocating parts of the task between the human and computer agent was already proposed in the seventies: adaptive automation (Chu & Rouse, 1979; Rouse, 1988; Parasuraman et al., 1992). Since mental effort investment must have a neurophysiological base, several researchers have investigated using brain-based measures for adapting the level of support provided by the technological subsystem.

Most famously, several studies used the 'engagement index', derived from the EEG, to initiate or withdraw task support. This strand of research was pioneered by Pope et al. (1995), who investigated several EEG frequency ratios, reflecting task engagement, to serve as a task support switch in a multitask environment mimicking a flight deck (the multi-attribute task battery; Comstock & Arnegard, 1992; Santiago et al., 2011). To validate their system and to compare between EEG-based triggers, these researchers looked at the system's behaviour. They reasoned that a valid EEG index has a functional relation with task engagement, which should lead to a relative stable, oscillating, pattern of providing and removing task support. The frequency ratio: beta/(alpha+theta), fitted this pattern best and was often used in subsequent research (Pope et al., 1995; Prinzel, 1997; Bailey et al., 2006; Freeman et al., 1999; Parasuraman, et al., 1999; Freeman et al., 2000; Prinzel et al, 2000; Berka et al., 2007). Overviews of research based on Pope's EEG engagement index can be found in: Prinzel et al. (2001), Scerbo et al. (2003), and Kaber & Prinzel (2006). Slightly different, Wilson & Russel (2007) showed that classifying between low and high workload during a simulated unmanned aerial task by feeding the electro-oculogram (EOG), EOG corrected EEG, and the electrocardiogram (ECG) into an Artificial Neural Network (ANN), could effectively be used for real-time task-support and improve task performance when compared to a random support schedule (see also, Christensen & Estepp, 2013). In addition, neurophysiologically based adaptive systems using cardiovascular activity have also been investigated. For example, Ting et al., 2010 combined the 0.10 Hz component of heart rate variability and the alpha and theta frequency band of the EEG through fuzzy logic modelling to predict the risk of a performance breakdown. This approach resulted in a performance improvement compared to an adaptive system triggered by a system error.

In essence, a user-adaptive system continuously updates a model of the user's mental state. As an illustration, the operator status model concept is provided in Figure 7.1 (Hoogeboom & Mulder, 2004; Mulder et al., 2008). In Figure 7.1, the status model receives input from two sources of information: physiology and task performance to infer the operator's current status.

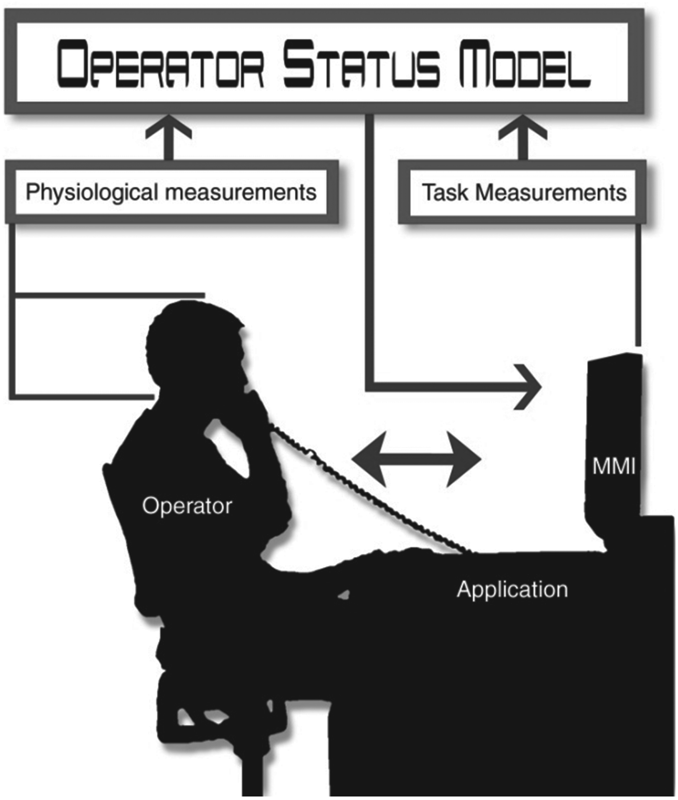



**Figure 7.1.** The Operator Status Model (OSM) concept. The OSM combines the operator's physiological information with task performance (taken from the active applications) and application states to derive the user's functional state, using generic dimensions like workload, visual task load and being occupied with a given task/application. The OSM information can subsequently be used by the application to change their HMI to optimise the direct information exchange (from Mulder et al., 2008).

Depending on the current status, the system may decide to act by providing or withdrawing aid. Alternatively, it could decide not to act if all parameters are within acceptable range. In addition, from adaptive automation literature it is clear that developing a dynamic user state model is crucial, but it is certainly not the only factor determining the effectiveness of an adaptive system. For example, an immediate coupling between inferred user state and adaptation of the system might be inappropriate under conditions of rapid changing task demands, as there will be a maximum adaptation frequency above which the human user is no longer able to efficiently keep up with the system (Morrisson et al., 1993; Hoogeboom & Mulder, 2004). This could necessitate the need to implement a ´deadband´ period when system-initiated adaptations are not possible. Then again, stability requirements must be balanced against responsiveness as a system that responds too late to workload increases is of little help to the human user or even leads to automation surprises because the link between the system change and its cause are not clear anymore (Morrison et al., 1993; Hoogeboom & Mulder, 2004).

A human-machine environment that might benefit from an adaptive support system is the driver-car system (e.g., Michon, 1992). Even though driving a car is a relative safe activity, traffic fatalities constitute one of the major causes of death throughout the world, simply because driving is such a common activity. The risk of running into an accident correlates to a number of conditions, both external (road, traffic, and environment; e.g., WHO, 2004) and internal (user state; e.g., Brookhuis & de Waard, 2010). Normally, an experienced driver adapts to these conditions, and these adaptations may be mapped onto one of three hierarchical levels of the driving task as described by Michon (1985; strategic level, manoeuvring level, and control level). On the strategic level, a driver may choose a different mode of transport, route, time of day, and so forth, to travel. On the manoeuvring level, a driver may choose not to overtake a car for example. On the more automatic control level, a driver might (subconsciously) increase safety margins, for example by keeping more distance with other traffic or slowing down. Through these behavioural adaptations, a driver keeps perceived risk, perceived workload or other related psychological constructs, at an optimum level or within an acceptable range (for an overview of theories of driving behaviour, see Lewis-Evans, 2012). An adaptive system that is aware of these internal and external conditions may aid the driver in adapting driving behaviour, for example by providing information, issuing warnings, switch off secondary tasks or even take over vehicle control if necessary (for examples, see Michon, 1985; Kohlmorgen et al., 2007; Dijksterhuis et al., 2012).

As stated, passive BCI is aimed at inferring user states from the electrical activities as measured on the scalp. For a preparatory study, we have investigated the feasibility of a

Common Spatial Pattern (CSP) based classification approach to discriminate between several levels of visuomotor workload elicited by manipulating speed and by providing driving performance targets in a driving simulator (Dijksterhuis et al., 2013). It was found that classification accuracies in high, gamma, frequency ranges, could discriminate between most workload levels with about 95% accuracy on average. Classification accuracies based on lower frequencies (below 20 Hz), were somewhat less, but still found to be in the 75-80% range.

The challenge for the current study was therefore to create and test a user-adaptive cruise control, which uses EEG-based workload classifications to slow down, speed up, or maintains the driving speed of the vehicle, thereby altering or maintaining the workload level. The BCI approach in the current study requires the classification model to differentiate between three levels of workload (low, comfortable, and high). Since the CSP technique is fundamentally limited to discriminating between two data classes, the multi-class classification requirement was met by applying multiple pair-wise classifications and a simple voting procedure (e.g., Friedman, 1996). In addition, we decided to equip the cruise control with a 'hard' driving performance loop in addition to a 'soft' BCI loop. In safety-critical tasks, such as driving a car, we feel that indications of risky driving (e.g., lane departures) should be prioritised over workload classifications. Decisions from the BCI loop were further subjected to a twenty-second deadband after each speed change decision, during which speed could not change (except in case of a collision with other traffic), which allowed the driver some time to get used to a new speed. Finally, since it is a well-established fact that EEG frequencies above 20 Hz are potentially contaminated by muscular activities (e.g., Whitham et al., 2007), we decided to apply two classification model versions separately in addition to driving the car without a BCI loop. The first version was trained on a frequency range between 5-20 Hz to minimise the influence of both EOG and muscular activity. For the second version, the system was allowed to choose between a classification model trained on the 20-45 Hz and the 55-80 Hz frequency band.

### 7.1.1 Hypothesis

As is common in BCI research, the classification models were trained on calibration data that were acquired directly before the application phase of the BCI system. During the calibration phase, EEG was sampled while the driver's drove above, at, or below their predetermined comfortable driving speed. The EEG data from the calibration phase were used for an offline classification study after the study was completed to assess the extent to which accuracies found during the preparatory study could be replicated.

Linking driving speed to EEG based workload classifications during the application phase of an experimental session was hypothesised to create a system that would keep the driver's workload at or oscillating around a comfortable level of workload. To test if the system would change speed to compensate for changes in the roadway environment, which reflected rural conditions, the number of curves that needed to be navigated per minute was manipulated by changing the curves' length. A high number of curves correspond to the high workload conditions, which should trigger the BCI system to set driving speed at a relative low level and vice versa. At this point, it should already be noted that the link between workload classifications and driving speed was delayed for several seconds for a number of participants. As will be argued in the discussion section, this does not imply that the results of this study are invalid, however, conclusions will be drawn carefully.

## 7.2 Method

### 7.2.1 Participants

In total, 18 participants (9 male) were recruited through poster announcements throughout the University of Groningen; they received 20 Euros for their time and effort upon completing the experiment. A substantial part of the participants were either Dutch or German students at this university. Ages ranged from 20 to 44 years (median = 23, interquartile range (IQR) = 4) and the participants had held their driving license from 1 to 25 years (median = 4, IQR = 4). The participants' reported average annual mileage over the last three years (or less if their driving license was obtained less than three years ago) ranged between 800 and 30,000 km (median = 1900, IQR = 5000) and self-reported total mileage ranged between 2000 and 800,000 km (median = 7000, IQR = 36,000). None of the participants reported on using prescribed drugs that might affect driving behaviour. The Ethical Committee of the Psychology Department of the University of Groningen has approved the study

### 7.2.2 Design and procedure

Upon arrival at the experimental site of the University of Groningen, a participant was informed in general words with respect to the experimental design, requested to sign an informed consent form, and asked to fill in a short questionnaire related to general demographic information and to their driving experience. Hereafter, the participant was given some time (ca. 5 min) to practice driving in the simulator, before the sensors were attached. During the experiment, a participant completed 14 short rides; an initial ride to establish a participants' comfortable driving speed, three calibration rides to sample the data that were used to train the BCI classification models, and nine application rides during which these BCI models were used to classify new, incoming EEG data. After each ride, the simulator slowed down and stopped the car before the participant was requested to fill in a short questionnaire. The three calibration rides, when driving speed was set below, at, or above the participant's comfortable speed (each lasted 140 s), followed directly after the initial ride and the order of these rides were balanced across participants. This resulted in a within subject-design with one repeated measures factor: driving speed (3) for the calibration phase. Then, two predictive BCI models (a low-frequency and a high-frequency model) were determined for further use during the application phase. During this phase, a participant drove on a road section consisting of either long, mid, or short curves, thereby creating three levels of lane keeping difficulty. In addition, driving speed control through the BCI system could be absent to create a control condition. In total, these manipulations resulted in a within-subject design consisting of two repeated measures factors: BCI-model (3) and curve length (3). The order of these nine conditions was balanced across participants according to a randomized Latin square. After finishing the last ride, the physiological sensors were removed, the participants were debriefed, and were paid 20 Euro upon leaving. In total, an experimental session lasted about two hours.

### 7.2.3 Simulator and driving environment

The experiment was carried out using the University of Groningen's StSoftware© driving simulator (http://www.stsoftware.nl), which consists of a fixed-base passenger car mock up with full controls. The computers that are dedicated to the simulator compute the road environment and traffic at 30Hz+, which are displayed on three 32-inch plasma screens and provide a total view of the driving environment of 210°.

For the experiment three 5.3 km and one 8 km two-lane road sections were prepared, consisting of two lanes (each 2.75 m wide), curving through rural scenery. The road surface was marked on the edges by a continuous line (0.20 m wide) and in the centre by a discontinuous (dashed) line (0.15 m wide). Outside the edges a soft shoulder was present and there were no objects in the direct vicinity of the road. Other traffic was not present in the participant's driving lane. However, oncoming traffic, travelling between 76 and 84 km/h, was generated with a random interval gap between 1 and 2 s, resulting in 40 passing private vehicles per minute on average. All sections consisted only of easy curves with a constant radius of 380 m. The curves' length differed per road section, creating a short, mid, and long curve condition.

The mid curve length was set to 125 m and the short and long curve length were set at respectively 75% and 125% of the mid curve length (94 m and 156 m). The 8 km road section consisted of 125 m curves only. Speed was set by the simulator during all rides, except for the initial ride during which the participants drove the simulator car (width: 1.60 m) in automatic gear mode. During this ride, participants drove on the long road section for three minutes and were asked to find and drive at a speed that felt most natural and comfortable in this situation while the speedometer was hidden from view. The average driving speed during the last two minutes of this initial ride was used as a reference point throughout the rest of the study.

### 7.2.4 Data acquisition

#### 7.2.4.1 Vehicle data

Driving speed and lateral position (LP) were sampled at 10 Hz. LP is defined as the difference in meters between the centre of the participant's car and the middle of the (right hand) driving lane. Positive LP values correspond to deviations towards the right hand shoulder and negative values correspond to deviations towards the left hand shoulder. During the nine application rides, sampled LP values were processed while driving and used to track how often any part of the vehicle swerved outside the driving lanes edges, by using a moving, overlapping time window of 20 seconds (which is equal to the system's deadband period), updated every 100 ms, and used to control driving speed in addition to a physiological loop and timing parameters (see Figure 7.2 for a full description of the system).

#### 7.2.4.2 Subjective ratings

After each ride, a rating on the one-dimensional Rating Scale Mental Effort (RSME; Zijlstra, 1993) was requested. The RSME ranges from 0 to 150, which can be used to rate experienced effort. In addition, several effort indications are visible alongside the scale, which may further guide the participant in marking the scale. Indications start with 'absolutely no effort' (RSME score of 2) and end with 'extreme effort' (RSME score of 112). In addition, participants were asked to fill in three questions with regard to the set driving speeds on 7-point likert scales. 1) "I felt comfortable with the vehicle´s speed." 2) " I felt that I could safely control the vehicle´s position on the road." 3) "I felt that the driving speed closely resembled the speed I would have chosen."

#### 7.2.4.2 Physiological measures

All physiological signals were sampled at 250 Hz. The EEG at all scalp locations in the International 10-20 system were measured using an electro-cap with 64 tin electrodes (at the following sites: FP1, FP2, Afz, F7, F5, Fz, F4, F8, T7, C5, C3, Cz, C4, T8, P7, P3, Pz, P4, P8, O1, and O2.) The amplifier was a REFA 8-72 (Twente Medical Systems International, Oldenzaal, The Netherlands). This device amplifies all channels against the average of all connected inputs (average reference). Registrations of the EEG were made using the build in functions of BCILAB. An Ag-AgCl electrode placed on the sternum served as the participant's ground. In addition, an Ag-AgCl electrode was attached to each mastoid. Impedances were kept below 20 kΩ for all electrodes.

### 7.2.5 The BCI approach

The BCI setup, both for offline and online purposes, was developed using a MATLAB toolbox for designing brain-computer interfaces: BCILAB 1.0 (Kothe; http://sccn.ucsd.edu/wiki/BCILAB).

#### 7.2.5.1 The calibration phase

Calibration data was acquired during three rides that immediately followed the initial ride. During these rides (140 s each), driving speed was either set below (75%), at, or above (125%) the previously determined comfortable speed, thereby creating a low, comfortable, and high visuomotor load condition. EEG data collected during the last 120 sec of each load-condition were divided into 1 s segments and used to train the BCI model. For this, the Common Spatial Pattern (CSP) technique, Fisher's Linear Discriminant Analysis (LDA), a voting procedure, and a model parameters search algorithm were applied.

#### 7.2.5.2 The Common Spatial Pattern

In general, the CSP technique improves discriminatory power of the data classifier by maximizing the contrast between two data classes. That is, CSP filters are determined in such a way that they maximize the variances of spatially filtered signals for one (workload) condition while minimizing them for the other (Blankertz, 2008). A CSP filter is a coefficient vector by which the original channels can be transformed. This results in new, spatially filtered, EEG signals (the CSP components), each of which is a linear combination of the original channels. The total number of filters and therefore, the number of components is equal to the number of original channels. The matrix of CSP filters is determined by solving a generalized eigen-value problem. The filter corresponding to the largest eigen-value yields a high variance signal in one condition, while producing a low variance signal in the other; and vice versa for the component corresponding to the smallest eigen-value. The CSP components are therefore ranked according to these eigen-values. To reduce dimensionality, either the first and last or the first two and last two filters in this sorted filter matrix were used for to transform the data in the current study (the number of used filters depended on an optimisation procedure, which will be described in more detail below).

#### 7.2.5.3 Fisher's Linear Discriminant Analysis

The next step in training a BCI model was to enter the log-transformed total variance of the CSP components of each of the 120 data segments (1 s each) per data class into a LDA. This analysis again transforms the data by determining the linear weights of the discriminant function that combines data points of the two data sets in such a way that maximizes the distance between them.

#### 7.2.5.4 Searching model parameters

For the current study, a comparison was made between a low and a high EEG frequency based classification model. The low range (5-20 Hz) was set to minimize contamination from electrical activity caused by eye movements and blinks in low frequencies and from muscular activities in high frequencies (e.g., Whitham et al., 2007). In addition, since higher frequencies proved to yield very good predictive results during a preparatory study (Dijksterhuis et al., 2013), either the 20-45 Hz or the 55-80 Hz frequency band was selected for use during the application phase. The selection criterion was classification accuracy as yielded by BCILAB's default optimization scheme for searching model parameters: a 5-fold block-wise cross-validation with five data segments as safety margin. This entails partitioning the data into 5 equal intervals, leaving a safety margin in between them, which was not used during the cross-validation procedure. A single interval was kept as validation data, while the classifier was trained using the other intervals. This was repeated until all intervals were used for validation exactly once. The prediction accuracy over these five folds represented the model's performance in a frequency band. Using a more elaborate validation scheme at this stage was considered but the idea was rejected to prevent unnecessary time delays before starting the application phase. Similarly, the program was allowed to select either one or two most contrasting component pairs from the CSP component matrix for both the low and high frequency BCI model. Again, BCILAB's default optimization scheme for searching model parameters was used.

#### 7.2.5.5 Voting

Since data classifiers are limited to classifying between two labelled data classes, a voting procedure was used to extend the model parameter optimisation procedure to three data classes (low, comfortable, and high workload). That is, for every optimisation fold, the classified workload level was determined by three pairwise classifications (high vs. low load, high vs. comfortable load, and comfortable vs. low load). Then, the probabilities that the classifiers assigned to each workload condition were summed and the condition to which the highest summed probability was assigned was the classification outcome.

#### 7.2.5.6 Calibration summary

In summary, two main BCI models were determined per participant, each consisting of three pairwise BCI sub-models. For the low frequency BCI model, a matrix of CSP components and linear weights for the LDA was determined per sub-model. These were used to classify the data. To allow for multi-class classification, a voting procedure was used. The performance of the multi-class classifier was retrieved through a mini cross-validation procedure. To optimize model parameters, the entire classification procedure was carried out twice, first using one and then using two component pairs, to determine which number of pairs resulted in the best accuracy. All model parameters: the number of included component pairs, the CSP component matrices for each data pair, the LDA weights, and frequency filter settings were then saved for use during the application phase. The parameter setting for the high frequency main BCI model was identical, except that the frequency band (20-45 Hz and 55-80 Hz) was also part of the parameter search procedure.

#### 7.2.5.7 Offline validation

To assess how well the trained BCI models classified the calibration data classes, results from an offline leave-one-out cross-validation (LOOCV) scheme will also be provided in the result section. This type of cross validation involves partitioning all data segments into one test-segment, while the remaining segments are used to train the classifier. This is repeated until all segments have been used to validate the classifier once.

#### 7.2.5.8 The application phase

The low and high frequency BCI classification models were applied to monitor workload while driving. That is, the model parameters as determined during the calibration phase were used to transform new data segments (1 s) into the probabilities that the new data segment resembles EEG data from the high, comfortable or low workload condition of the calibration phase. In addition, these classification probabilities, which were updated every 100 ms, were combined and smoothed using an exponentially weighted moving average (EWMA; alpha = 0.01). The data class to which the highest weighted probability was assigned, constituted the classification outcome. Finally, these outcomes were used as a control signal to set the car's driving speed in addition to several vehicle and timing parameters. For example, a classification 'high workload' was used to advise the decision module of the simulator software to set driving speed at 75% of the current driving speed.

### 7.2.6 Speed change decisions – the system architecture

As mentioned before, in addition to the BCI system, vehicle parameters were also involved in determining the car's driving speed. This effectively created two feedback loops: a 'hard' safety loop based on lane keeping performance and a 'soft' physiological loop based on workload classifications. The safety loop advised the decision module that lateral control was either safe or unsafe, while the physiological loop advised that workload was either high, comfortable, or low. In case lateral control was unsafe, driving speed was reduced, regardless of workload classification. Lateral control was defined as the percentage of driving time that any part of the car swerved outside the driving lane during the last 20 s and this percentage was updated at 10 Hz. The criterion for unsafe lateral control was set at 40%. The main reason for implementing a control loop based on driving performance was to prevent that misclassifications could speed up the car to the point that loss of vehicle control can be expected. In case the safety loop advised safe vehicle control, a high, comfortable, or low workload classification could lead to a decrease, maintenance, or increase of the current driving speed. In addition, the decision module locked current driving speed for 20 seconds after a speed change had been initiated. Next, collisions needed to be handled. Although a lateral performance loop was implemented to prevent such unfortunate events, it did not actively intervene with steering to prevent a participant from colliding with oncoming traffic. In case of a collision, the simulator car did not actually crash, but continued to drive at a reduced speed after it was positioned back on the driving lane by the simulator software. Finally, an absolute minimum and maximum for a new driving speed was set (40 km/h and 160 km/h), which was equal for all participants. That is, a new, lower or higher driving speed was not allowed if current driving speed was already outside this range, but a new speed could in fact be set outside these limits (e.g., a speed change from 50 to 37.5 km/h). The absolute speed limits were set to prevent the system from setting new speeds that are clearly outside the range of normal driving speeds given the rural road environment. In summary, 5 factors were taken into account by the decision module to set driving speed (in order of priority): an absolute minimum and maximum driving speed, the occurrence of a collision, the speed change deadband, the vehicle parameter feedback loop, and the physiological feedback loop. The entire system is schematically presented in Figure 7.2.

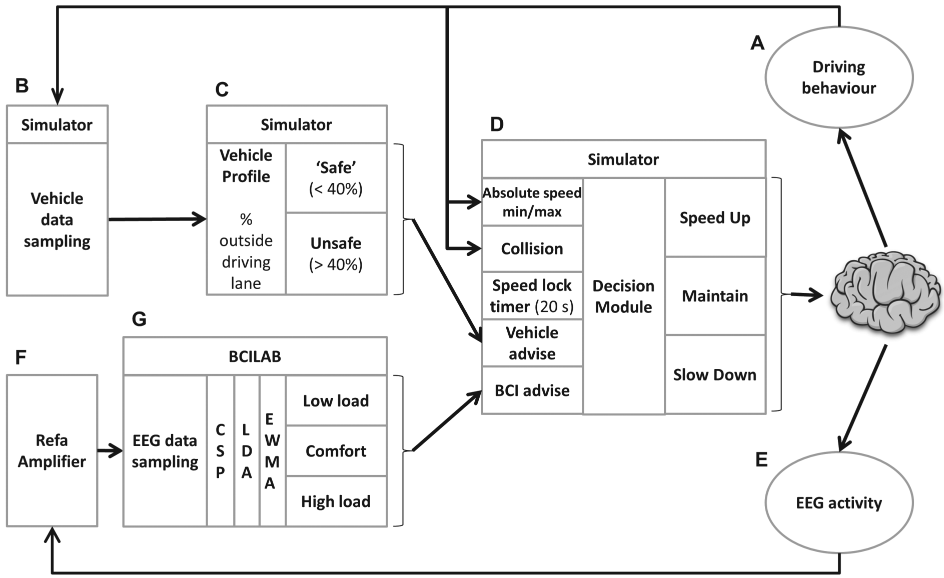



**Figure 7.2.** An overview of the system's architecture. Five factors could possibly initiate or prevent a change in driving speed (see (D)). 1) Driving at very low or very high speeds. 2) The occurrence of a collision. 3) The speed change 'deadband', 4) The vehicle's advice with respect to lane keeping behaviour. 5) The BCI advise with respect to the level of current workload. These factors are listed in order of priority and speed changes always occurred as 25% of the current driving speed. A participant's lateral position (LP) on the driving lane (A) was sampled at 10 Hz (B) which was used to track the percentage that any part of the vehicle swerved outside the driving lane in the last 20 s (C). If over 40%, the (C) module advised the decision module (D) to decrease speed. The decision module would then set speed at 75% of the current driving speed if 1) current driving speed was below 160 km/h and over 40 km/h and 2) the last speed change had occurred more than 20 s before. A speed change affects driving behaviour (A) and EEG activity (B). EEG activity was amplified (F) and sampled at 250 Hz (G), before the predetermined Common Spatial Pattern (CSP) and Fisher's Linear Discriminant (LDA) coefficients per data class pair were combined, using a voting procedure, to assign the probability that a new data segment (one second) belonged to low, comfortable or high workload. In addition, these probabilities were updated at 10 Hz and smoothed using an exponentially weighted moving average (EWMA; alpha = 0.01), which resulted in an BCI advise to the decision module (D). Low, comfortable, and high load resulted in a 'speed up', 'maintain speed', and 'slow down' advise respectively, which was used by the decision module to change speed if all other factors allowed for it. For example, the occurrence of a collision or a vehicle advise to slow down had priority over the BCI loop. The occurrence of a collision resulted in a 25% decrease of set driving speed.

### 7.2.7 Statistical testing

For all tested variables, repeated measures ANOVA's were performed using IBM SPSS Statistics 20 to test the differences between the various experimental conditions. In case Mauchly's test for sphericity showed a significant effect, the degrees of freedom were adapted according the Greenhouse-Geisser method. Polynomial contrasts and Bonferroni corrected pairwise comparisons were used to explore the nature of significant main effects.

## 7.3 Results

### 7.3.1 Classification delays – Technical difficulties

After the experiment was completed, it was discovered that for a number of rides, it is likely that the weighted classifications that were send from the PC running BCILAB to the PC running the simulator software, may have had a time delay. This was discovered after all classifications were reconstructed by carrying out a pseudo-online classification procedure, in which the EEG data that was recorded during the experiment was again classified and compared to the classifications as recorded and used by the simulator software. Since the data and procedure from the online and the pseudo-online study are identical, the classifications over time should also be identical. This was not the case for substantial portion of the data. However, this did not affect result from the calibration phase.

### 7.3.2 The calibration phase

#### 7.3.2.1 Vehicle and subjective data

The results related to vehicle parameters and subjective ratings are shown in Figure 7.3 and their test results are listed in Table 7.1. As can be read from Table 7.1, all variables shown in Figure 7.3 changed significantly as a function of driving speed. To start with, lateral position on the driving lane (see Figure 7.3A) shifts in an almost straight line from 0.10 m towards the right hand shoulder during the fast ride to 0.03 m during the slow ride (polynomial contrasts showed a linear trend: $F(1,17) = 6.36$, $p = 0.022$, $\eta_p^2 = 0.27$). Given that the dimensions of the vehicle and driving lane allowed for 0.58 m of swerving margin on both sides of the vehicle, this shows that drivers stayed well within their driving lane on average.

However, occasionally the drivers did allow the car to swerve outside the driving lane (see Figure 7.3C and Figure 7.3F). The vehicle drifted (partly) onto the opposite driving lane for about 4% during the fast ride (see Figure 7.1C) but this percentage was lower for the two slower rides (quadratic trend: $F(1,17) = 15.64$, $p = 0.001$, $\eta_p^2 = 0.479$). The time that any part of the car strayed over the right lane edge was about 8% of the total time, but as can be seen in Figure 7.1F, this percentage decreased strongly as speed decreased (linear trend: $F(1,17) = 14.93$, $p = 0.001$, $\eta_p^2 = 0.468$). The standard deviation of the lateral position (SDLP), reflecting swerving behaviour, decreased as driving speed decreased (linear trend: $F(1,17) = 46.00$, $p < 0.001$, $\eta_p^2 = 0.479$). However, this increase was slightly more pronounced from fast to comfortable speeds, creating a quadratic trend as well ($F(1,17) = 5.39$, $p =0.033$, $\eta_p^2 = 0.241$). Subjective effort shows a clear linear trend ($F(1,17) = 37.93$, $p <0.001$, $\eta_p^2 = 0.691$), as drivers rated each faster speed category about 16 RSME point higher than the slower one. Effort ratings range from a little over 'a little effort' (RSME = 28) for the slow speed condition to a

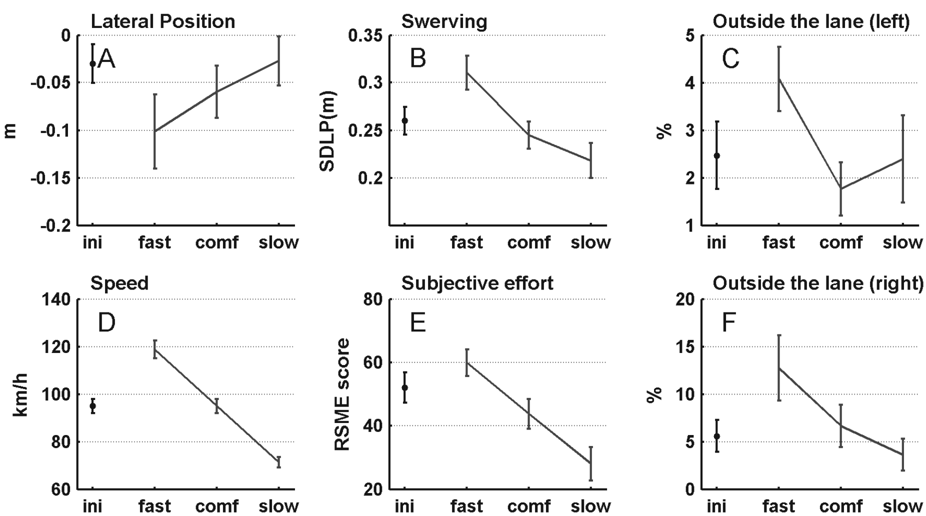



**Figure 7.3.** Vehicle parameters and subjective ratings as a function of driving speed condition. (A) Mean Lateral Position (LP). LP values represent the middle of the car (car width = 1.60 m) in relation to the middle of the right (driving) lane (width = 2.75 m);negative values reflect a deviation towards the right hand shoulder. (B) Standard Deviation of the Lateral Position (SDLP) reflects swerving behaviour. (C) The percentage of the total time that any part of the vehicle strayed left of the driving lane. (D) Mean Driving Speed. (E) Rating Scale Mental Effort (RSME). Maximum score for mental effort is 150. (F) The percentage of the total time that any part of the vehicle strayed right of the driving lane. On the x-axes, values for the initial ride (ini) are shown in addition to three driving speeds that were set, relative to the individual's preferred driving speed established during the initial ride; either 125%, 100%, or 75% of the preferred, comfortable speed. These driving speeds reflect high, comfortable, and low visuomotor workload. Error bars represent the standard error. n=18.

little over 'rather much effort' (RSME = 60) for the fast speed condition. Finally, Figure 7.1D shows driving speeds. The average speed chosen by the participants during the initial ride is 95 km/h, ranging from 71 to 116 km/h. Not surprisingly, since driving speeds during the other calibration rides were set by the driving simulator, speed shows a large significant effect (see Table 7.1)

**Table 7.1.** Univariate test results for vehicle parameters and subjective effort ratings (Figure 7.3). LP = lateral position, SDLP = standard deviation lateral position. Degrees of freedom were Greenhouse-Geisser adjusted if Mauchly's test of sphericity showed that sphericity could not be assumed.

| **Effects of set driving speed on vehicle parameters and subjective effort** | | | | |
|---|---|---|---|---|
| *Variable* | | **F(df1,df2)** | **p** | $\eta_p^2$ |
| **Lateral Position (LP)** | Fig 7.3A | 4.90 (1.5,28.2) | 0.020 | 0.224 |
| **SDLP** | Fig 7.3B | 32.31 (2,34) | <0.001 | 0.655 |
| **Outside Lane (left)** | Fig 7.3C | 5.95 (1.4,24.6) | 0.013 | 0.259 |
| **Speed** | Fig 7.3D | 1037.17 (1.5,17.0) | <0.001 | 0.984 |
| **Subjective effort** | Fig 7.3E | 29.35 (1.5,25.8) | <0.001 | 0.633 |
| **Outside Lane (right)** | Fig 7.3F | 13.33 (1.0,22.1) | 0.001 | 0.440 |

#### 7.3.2.2 Model training

As described in the method section, the BCI computer program searched between several model parameters during the calibration phase to optimise classification performance. To start with, the optimisation procedure could choose from using one or two CSP filters to transform the original EEG channels and the most accurate BCI model would be selected for further using during the application phase. As summarised in Table 7.2, BCI models using two filter pairs usually yielded superior classification accuracies compared to BCI models that used just one pair, as these were selected about three times more often for both the low and high frequency model. In case of optimising the high-frequency model, when the optimisation procedure was allowed to choose between the 20-45 Hz and 55-80 Hz frequency band to classify the data, the higher frequency band resulted in a better accuracy for most participants (15 vs. 3).

Model selection was based on a mini cross-validation that was carried out between the calibration and application phase of each experimental session. However, a more extensive procedure is more common when validating how well a classification model classifies the data. For this purpose, a leave-one-out cross-validation procedure was carried out, using the same model parameters that were selected through the optimisation procedure. As can be read from Table 7.3, the average classification accuracy when using the high frequency

**Table 7.2.** An overview of selected model parameters. For the low frequency version, only the number of filter pairs are relevant in the optimisation procedure. For the high frequency version, frequency band was also included. Model selection was based on classification accuracies that resulted from a mini cross validation procedure during the experimental sessions. n=18.

| Number of filter pairs | | 1 | 2 |
|---|---|---|---|
| **Low frequency model** | **5-20 Hz** | 5 | 13 |
| **High frequency model 1** | **20-45 Hz** | 1 | 2 |
| **High frequency model 2** | **55-80 Hz** | 3 | 12 |

**Table 7.3.** Average workload classification accuracies for the calibration data. The percentages are the result of a post-experimental, exhaustive leave-one-out cross validation procedure (119 one-second segments for each target data class). The model parameters (number of component pairs and frequency band) that were used to produce the accuracies are identical to those selected during the experiment. Per one-second segment, the workload classification is determined by combining three pairwise classifications (high vs. low load, high vs. comfortable load, and comfortable vs. low load). The probabilities that the classifiers assign to each class are summed. The data class, to which the highest summed probability is assigned, is the winner of this voting procedure. Note that chance level is 33% given three data classes. N=18. Low = low frequency BCI model (5-20 Hz). High = high frequency BCI model (20-45 Hz or 55-80 Hz depending on the parameters search outcome).

| **'Low'** | | Overall Accuracy **68%** | | | **'High'** | | Overall Accuracy **90%** | | |
|---|---|---|---|---|---|---|---|---|---|
| | | Prediction | | | | | Prediction | | |
| | (%) | High | Com | Low | | (%) | High | Com | Low |
| Target | High | **71** | 18 | 12 | Target | High | **92** | 5 | 3 |
| | Com | 14 | **65** | 20 | | Com | 2 | **89** | 9 |
| | Low | 14 | 18 | **67** | | Low | 2 | 7 | **90** |

models is 90%, which is 22% higher than the average low frequency models' output. The accuracies differ somewhat between the data classes (high, comfortable, and low visuomotor load), as the average accuracy for the comfortable workload level is lower than the other two classes, and the high visuomotor load was best classified.

To get an indication of the consistency of classification accuracies across participants, cumulative accuracies are displayed in Figure 7.4. For instance, for 16 out of 18 participants, classification accuracy is 80% or better when using the high frequency BCI model. In the low frequency range, this is true for only three participants. Similarly, one participant scored better than 95% when classifications were based on the low frequency band, while five participants scored higher than this when classifications were based on the higher frequency bands.

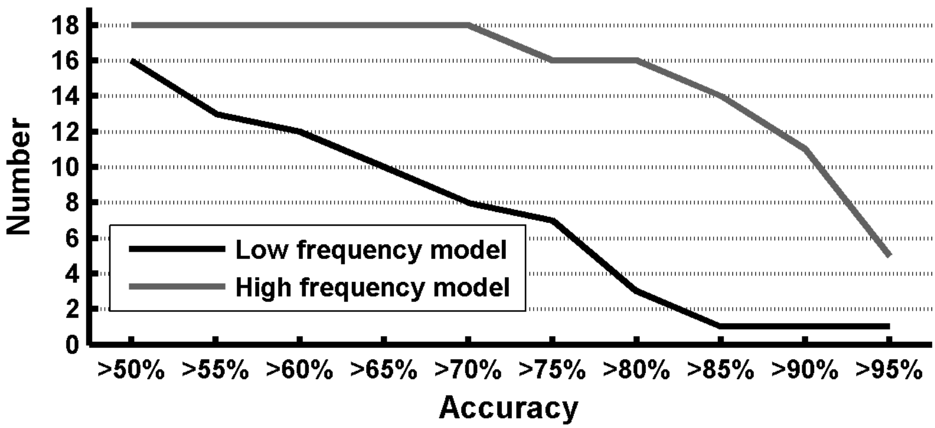



**Figure 7.4.** The cumulative frequencies of classification accuracies. The number of participants for whom a particular accuracy was found or better (max = 18 participants) in the accuracy categories on the x-axes, for both the low frequency BCI model and high frequency BCI model which were determined during the calibration phase. Note that chance level is 33% given three data classes.

### 7.3.3 The application phase

During the application phase, the low and high frequency BCI models that were trained and selected at the end of the calibration phase were used to classify new, incoming, EEG data while the participants were driving. These workload classifications (low, comfortable, or high) were subsequently used to change or maintain driving speed. Since several other factors also contributed to the final speed decision (namely, absolute driving speed, the occurrence of a collision, a speed change deadband, and indications of worsened lane keeping behaviour), the behaviour of the total system will first be described.

As explained above, the system responded to workload classifications that are now suspected to have had time delays. Mostly, these were shorter than 1 s, but occasionally grew to about 5 s. Given that the system responsiveness was not very fast due to the 20 s deadband, these delays may have been relatively harmless for the system's behaviour from the perspective of the driver. Therefore, examples of time-on-task results are still provided, representing several categories of system behaviour.

#### 7.3.3.1 System behaviour – setting driving speed

The total driving time over all participants, during which the brain-car interface was monitoring workload (either the low or high frequency BCI model), added up to 6 h, 42 min, and 10 s. During this time, the system changed speed 726 times, which is once every 33 seconds on average and 13 seconds on average when subtracting the dead-band period (20 s) during which speed was only allowed to change in case of a collision. Please note that these average periods are just a rough indication of the system's speed setting behaviour, since the driving speed was set below the absolute minimum speed criterion (below 40 km/h) for about a third of the total driving time, which disabled any further speed decreases. In addition, there were 20 collisions, which also happened in a dead-band period, and there were a number of speed changes within 20 s before the end of an experimental condition. In total, speed was increased 365 times, which only happened if the BCI loop advised this while all other factors allowed for it (see Figure 3D for details). Speed was reduced 361 times: 20 times as a result of the occurrence of a collision, 108 times because the threshold for unsafe lane keeping behaviour was passed, and 233 times because the BCI loop indicated high workload.

The system behaviour, as seen from the driver's perspective (i.e., the speed changes as they occurred over time), showed a large amount of inter-individual differences. Therefore, making general statements that are true for everyone or most is difficult. However, some general differences in the way the system behaved may be discerned. A couple of examples will be described below.

**Example 1. Stable system, driving at minimum speed**

To start with, during a number of rides, driving speed was stable for large parts of the ride. This behaviour could be caused in several ways. For example, during a number of rides, the aforementioned speed threshold level below which changing to a lower speed was not allowed was reached during some point. This pattern occurred during 18 out of a total of 105 conditions (18 subjects * 6 conditions during which the BCI feedback loop was active, minus 3 missing data points). When this happened, driving speed typically remained very low for large portions of a ride (65% on average). Figure 7.5 illustrates such a case. New EEG data was consistently classified as 'high workload' (Figure 7.5B) and, taking into account the dead-band period of 20 s, this caused driving speed to drop three times in the beginning of the ride (Figure 7.5A and Figure 7.5B). Subsequent speed decreases were not allowed and therefore no more decisions to change speed were made for the remainder of the ride.

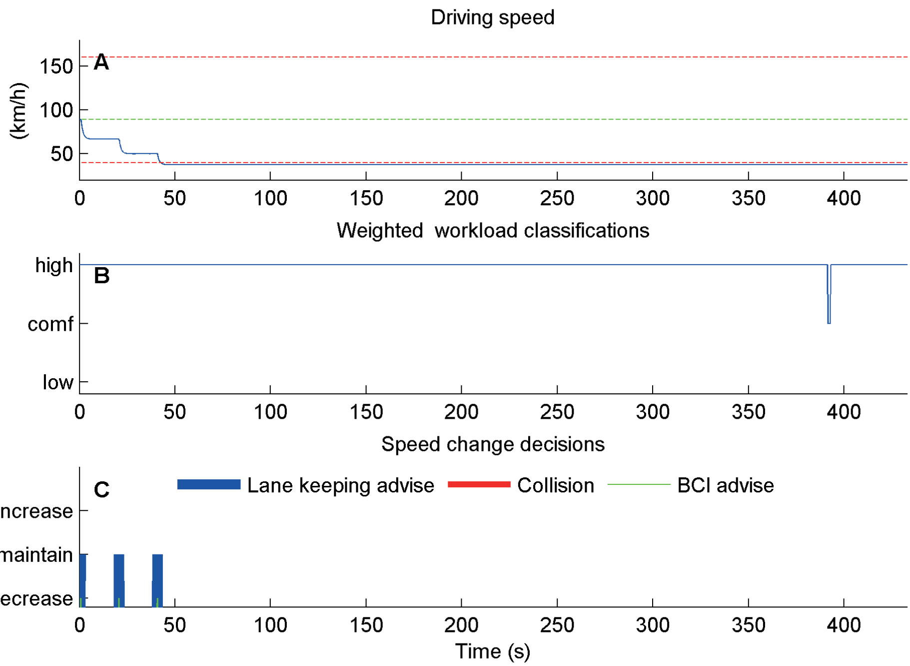



**Figure 7.5.** An illustration of the system's behaviour over time for subject 11, condition 7 (applying the high model / long curved road section). (A) Driving speed. The horizontal red dashed lines represent the absolute minimum and maximum speed (40 and 160 km/h). The horizontal green dashed line represent the participant's comfortable driving speed as established during the initial ride (102 km/h). (B) The weighted workload classifications that were used to advise driving speed. (C) The final speed decision, taking into account all decision factors (absolute min/max, collisions, dead-band period, proportion of driving time outside the driving lane, and the EEG advice. In general, the BCI advice can be either to increase, maintain, or decrease speed. The lane keeping advice can only be to maintain ('safe' lane keeping behaviour), or decrease ('unsafe') driving speed. The occurrence of a collision immediately results in a speed decrease, which did not happen during this ride.

## Example 2. Stable system, comfortable level

Another example of a stable system, which occurred a number of times, is illustrated in Figure 7.6. In this case, new EEG data was always classified as being the result of comfortable workload level. Therefore, it is advised to maintain the current, comfortable, driving speed for the entire ride. The speed decrease at about 162 s was the result of the participant drifting over the lane edges for over 12 s during the preceding 20 s. However, this did not affect workload classifications.

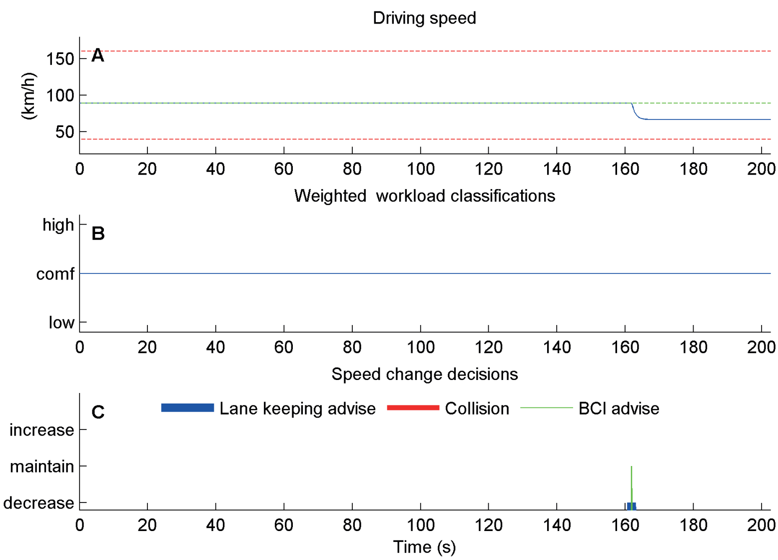



**Figure 7.6.** An illustration of the system's behaviour over time for subject 10, condition 8 (applying the high model / mid curved road section).

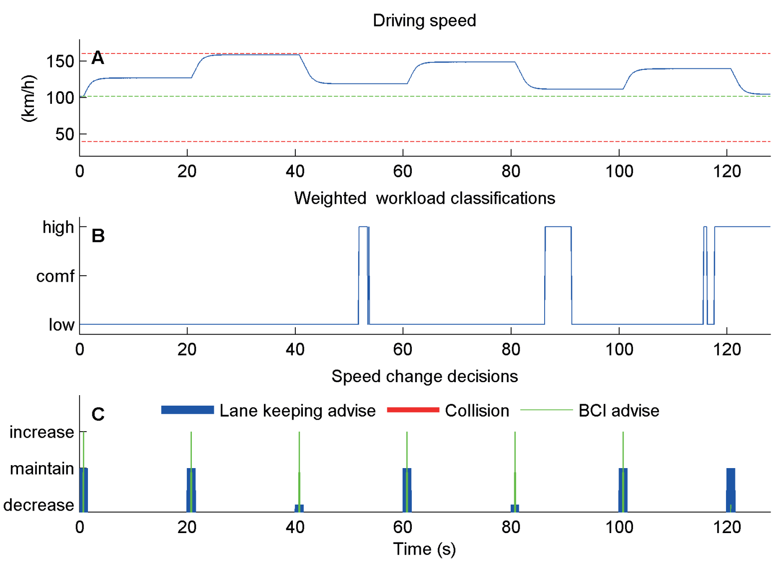



**Figure 7.7.** An illustration of the system's behaviour over time for subject 7, condition 5 (applying the low frequency model / mid curved road section).

**Example 3. Oscillating system, unsafe lane keeping**

Another main category of system behaviour may be described as 'oscillating speeds'. Again, several underlying advise patterns may have caused this type of system behaviour. For example, for a number of rides, incoming EEG data was largely labelled as 'low workload' (see Figure 7.7). Therefore, it mostly advised to increase speed. Whilst driving at high speeds, participants regularly crossed the threshold for unsafe lane keeping behaviour, which was prioritised over the BCI advice. The driving speed was subsequently decreased and locked for the next 20 s, after which lane keeping behaviour had recovered and the BCI loop again advised to increase speed and so forth.

**Example 4. Oscillating system, longer periods of high and low workload**

Another example of an oscillating system is given in Figure 7.8. In this example, vehicle parameters were not involved in changing speed at all (see Figure 7.8C). During the first third of this ride (up until about 110 s), workload classifications are mostly 'high', causing driving speed to drop several times, even below the criterion after which more speed decreases are not allowed. However, at that point, workload classifications shift to 'low' for about 80 s, leading to multiple speed increases. This pattern was then repeated for the remainder of the ride.

**Example 5. Variable patterns**

Finally, there are a number of rides for which, a stable or oscillating pattern do not really apply. For instance, driving speed could be relative stable for one part of a ride, but shows unstable, non-oscillating pattern for the other. During the course of the first half of the ride that is illustrated in Figure 7.9, driving speed remained at or around the comfortable driving speed, and peaked at about 110 s when the participant could not prevent a collision with oncoming traffic. Hereafter, workload classification started to go to 'high', causing four subsequent drops in driving speed to under 40 km/h. Not until the end of the ride, did workload classifications change again.

7.3.3.2 Subjective ratings.

Figure 7.10 displays the effects of model type and curve length on the subjective data. To start with, curve length did not affect the scores on any of the variables (see Table 7.4). However, model type did have a main effect on all variables. Upon taking a closer look in all subfigures of Figure 7.10, it appears that in conditions without BCI feedback, subjects rated subjective

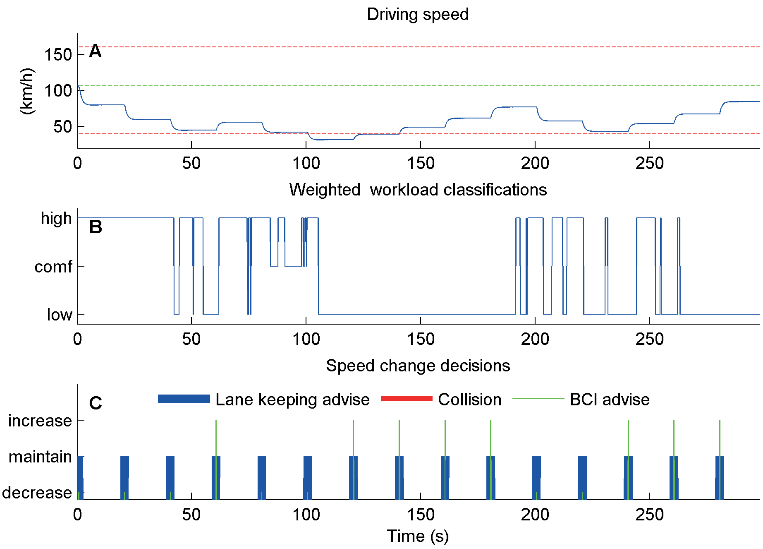



**Figure 7.8.** An illustration of the system's behaviour over time for subject 17, condition 5 (applying the low frequency model / mid curved road section).

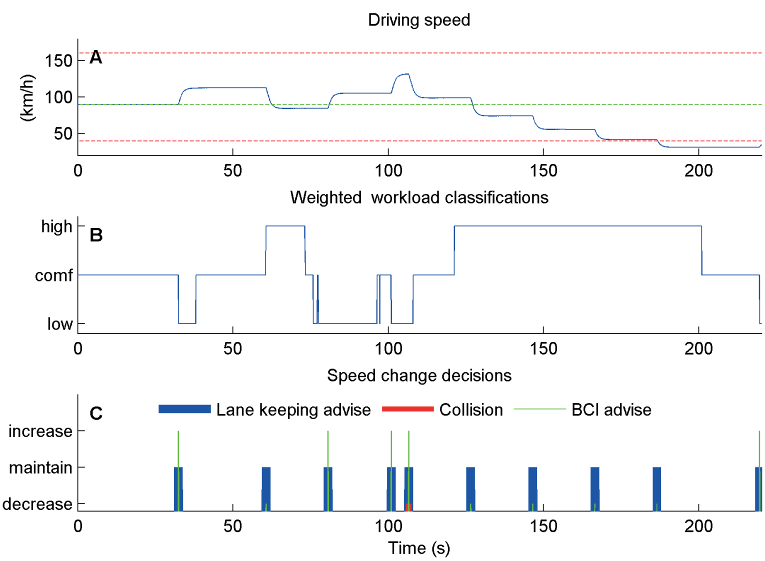



**Figure 7.9.** An illustration of the system's behaviour over time for subject 18, condition 4 (applying the low frequency model / short curved road section).

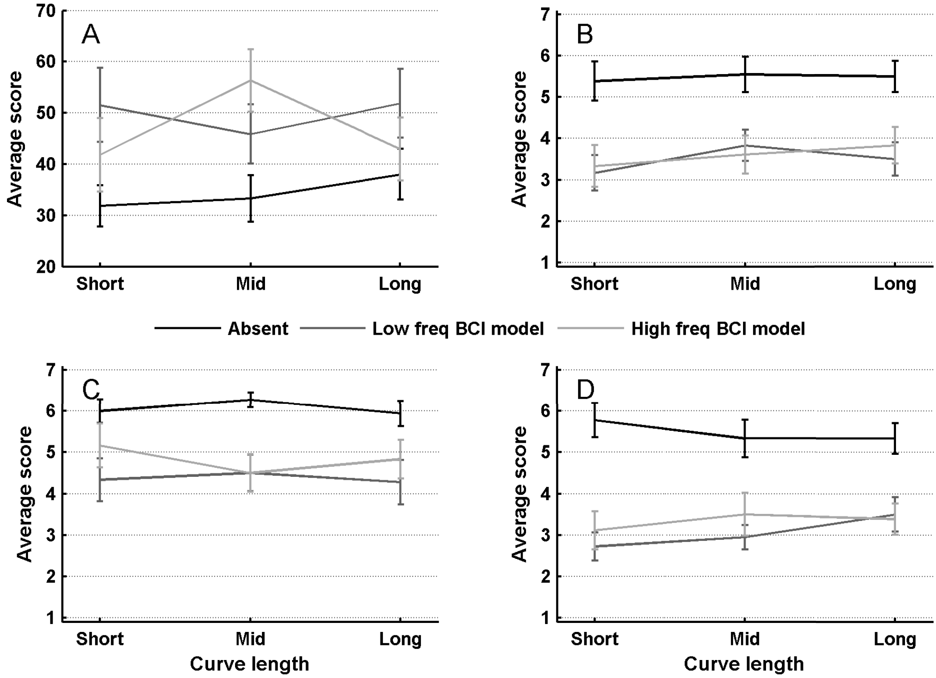



**Figure 7.10.** Average subjective data for all experimental conditions. (A) Subjective effort (RSME). The RSME scale ranges from 0 to 150. (B-D) The average scores on 7 point likert-scales. (B) 'I felt comfortable with the vehicle's speed.' (C) I felt that I could safely control the vehicle's position on the road.' (D) 'I felt that the driving speed closely resembled the speed I would have chosen.' The error bars represent the standard error.

effort lower (Figure 7.10A), felt more comfortable with the driving speed (Figure 7.10B), felt that they could control the vehicles position on the road more safely (Figure 7.10C), and felt that driving speed resembled the speed that they would have chosen themselves more closely (Figure 7.10D). This is confirmed by pairwise comparisons (Bonferroni corrected), where the difference between BCI-absent and the other two are significant but the difference between the low and high frequency model does not differ significantly, for any of the variables.

## 7.4 Discussion and conclusions

The aim of this study was to investigate a proto-typical adaptive system aimed at keeping visuomotor workload within a normal range in a driving simulator. Per participant, two versions of a BCI classification model were trained on calibration data. During the application

**Table 7.4.** Univariate test results for subjective effort ratings. RSME = rating scale mental effort. Significant effects ($p < 0.05$) are shown in bold. Mauchly's test for sphericity was not significant for any of the factors, so sphericity was assumed.

| **Subjective data** | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | RSME | | | Q1 (Comfortable?) | | | Q2 (Safely?) | | | Q3 (Resemble?) | | |
| *Effect* | F | p | $\eta_p^2$ | F | P | $\eta_p^2$ | F | p | $\eta_p^2$ | F | p | $\eta_p^2$ |
| **Model (M)** | 9.01 | 0.001 | 0.35 | 18.12 | <0.001 | 0.52 | 8.31 | 0.001 | 0.33 | 23.63 | <0.001 | 0.58 |
| **Curve length (L)** | 0.60 | 0.554 | 0.03 | 1.12 | 0.338 | 0.06 | 0.01 | 0.862 | 0.09 | 0.26 | 0.772 | 0.02 |
| **M x L** | 2.31 | 0.066 | 0.12 | 0.18 | 0.947 | 0.01 | 0.73 | 0.573 | 0.04 | 1.07 | 0.379 | 0.06 |

phase of the experiment, these models were used to classify workload levels from incoming EEG data which were then used to alter driving speed. The effects of the various experimental manipulations, both during the calibration and application phase of the study, on driving behaviour, subjective ratings, workload classifications, and driving speed will be discussed.

To start with, speed was set below, at, and above the comfortable driving speed during the calibration phase, which was established during the initial ride. The effect of driving speed on subjective data and vehicle parameters were as expected. That is, the standard deviation of the lateral position (SDLP), indicating swerving behaviour, increased as speed increased, which has been found in other simulator studies as well (e.g., Dijksterhuis et al., 2013; Peng et al., 2012). The average lateral position on the road shifted from the centre of the driving lane towards the right hand shoulder as speed increased. The reason that drivers increase their lateral distance from oncoming traffic may be related to maintaining safety margins by the driver (Summula, 2003). As speed increases, a driver has less time to carry out small steering wheel corrections to keep the vehicle on the lane. Increasing the distance to oncoming traffic is a way to compensate. As a consequence of driving towards the right and increased swerving behaviour, the time that the vehicle drifted outside the lane's edges also increased. Also, higher speeds were rated as being more effortful (similar to Dijksterhuis et al., 2013), which indicates that workload level was successfully manipulated.

After the calibration rides were completed, the computer trained several classification models and selected two of them using a parameter optimisation procedure. To validate how well the trained models classify the three data classes, an off-line leave-one-out cross validation procedure was carried out, using the model parameters selected during the experimental sessions. As it turned out, the classification model accurately assigned 90% of the 1 s segments to the correct data class for the high-frequency model version on average. The low-frequency model accuracy was remarkably lower at 68%. A further confirmation that higher frequencies tend to yield higher classification accuracies is the fact that the optimisation procedure in most cases (15 out of 18) selected the highest frequency range for the high frequency BCI model. This replicates results from a preparatory off-line classification study (Dijksterhuis et al., 2013).

Regretfully, the interactive system during the application phase responded to workload classifications that are now suspected to have had time delays. Mostly, these were shorter than 1 s, but occasionally grew to about 5 s. An immediate link between workload classifications and driving speed can therefore not be assumed when interpreting the results of the application phase, even if all other factors allowed for it. Despite this technical problem, we argue that these delays may have had a relative low impact on the driving experience, because the system's responsiveness was already relative slow due to its dead-band. As a consequence, the results did not lose all validity, even if it limits the conclusions we can draw from this study.

The subjective effort ratings from the application phase suggest that the curve length manipulation did not result in the intended visuomotor workload levels (see Figure 7.10), nor did any of the other subjective ratings differ as a result of this manipulation. A priori, it was wrongfully assumed that a 25% change per curve length level during the application phase would evoke changes in subjective effort ratings that were similar to the 25% change in driving speed during the calibration phase. The absence of this effect may be due to the fact that a speed change does not only change the number of curves that needs to be navigated per minute, but also decreases the available time for making steering wheel correction, e.g., before a lane departure occurs. In hindsight, it may be concluded that this latter factor was dominant for the effort ratings.

Next, participants experienced the two versions of the classification model (high and low frequency) very similarly and rather negatively compared to the ride without a BCI loop. This indicates that participants found driving with a BCI loop more effortful, did not really feel comfortable driving with a BCI loop, nor felt that the system resembled their own speed choices. Partly this may have been caused by the fact that, unlike existing cruise controls, the participants could not turn off the device. Another reason for these unfavourable scores may be related to the specific system's behaviour, which allowed for a large range of speeds, which at times was clearly under or above the driver's comfortable speed and will be discussed in more detail below.

As it turned out, system behaviour varied immensely from ride to ride as a result of the interacting decision factors: the absolute minimum and maximum driving speed, the occurrence of a collision, the dead-band period, the vehicle's advice, and the BCI advice. However, upon closer inspection, two major distinctions between system behaviour can be made. First, a relative stable system could be the result of stable comfortable workload classifications combined with 'safe' lateral control, resulting in speeds at or close to the comfortable driving speed. Second, for a number of rides, the BCI loop continuously advised to decrease speed. This would cause the simulator to slow down to the absolute minimum driving speed, since none of the other decision factors could trigger a speed increase. Although this resulted in driving at slow speeds, it did not result in unsafe driving. On the other hand, during a number of rides, the BCI loop continuously advised the decision module to increase driving speed. This resulted in an oscillating system, since driving at high speeds almost unavoidably triggered the lateral vehicle control loop to advise to slow down the vehicle. Typically, these speed changes occurred every 20 s, immediately after the dead-band period had passed. However, oscillating speeds were also observed as a result of oscillating workload classifications, although they were not always directly linked. That is, several speed decreases could be observed before the majority of workload classifications changed to 'low workload' and vice versa.

Stable workload classifications at the comfortable level was expected for some rides since the mid curve length conditions of the application phase was identical to the conditions of the calibration phase. In other words, continuous low and high workload classifications, despite substantial changes of the driving speed, were not expected. It is possible that the link between driving demand and the individual's workload levels altered during the course of an experimental session, since time-on-task related effects may influence workload regardless of task demand (see e.g., Hockey, 2003). However this seems insufficient to explain these misclassifications. Many other possible sources may have contributed to the degradation of BCI classification performance over time, which is usually referred to as non-stationarity of the BCI features or the underlying EEG signals (e.g., Shenoy et al., 2006; Lotte et al., 2007; Krauledat, 2008). In other fields of human-machine interaction, time and context dependency are referred to as issues of psychophysiological reliability (e.g., Brookhuis & de Waard, 1993; Prinzel, 2001; Hoogeboom & Mulder, 2003; Fairclough, 2009; Hockey et al., 2009; Mulder et al., 2009).

To illustrate one case of non-stationarity, during an experimental session, one of the peripheral EEG electrodes started to display large amplitude activities after two calibration rides, and remained to show this activity for the remainder of the experimental session. What may have caused this noisy signal is not entirely clear, however, since the CSP technique used in this study assigns the highest weight to the most discriminating EEG channels, it classified all incoming data into the data class that was last experienced during the calibration ride. This caused the system to be biased towards 'high workload' classifications. This also illustrates the vulnerability of using data-driven machine learning algorithms. The CSP uses all information it receives to distinguish between one data class and the other. In this way, it may have been accidentally trained on non-task-related physiological factors, such as muscular activities.

Yet another explanation for misclassifications may be related to the range of speeds that the classification models used as input. During the calibration phase, speeds were set at 75%, 100%, and 125% of the initial driving speed. During the application phase, speed could increase or decrease over a much broader range (from about 40 km/h to 160 km/h). This approach was preferred over designing an interactive system that could only vary between three driving speeds, and over expanding the number of calibration speeds to a high number of data classes. As a consequence, it was assumed that driving at speeds below the low-speed condition of the calibration phase would naturally lead to a low-workload classification as well and vice versa. However, this might not be necessarily the case. For example, if a change in visuomotor workload qualitatively changes the underlying neural patterns (e.g., different areas of the cortex become important as workload increases) then the CSP filter as determined during the calibration phase is inaccurate, since it assigned the most extreme linear weights to other scalp locations.

From this study it is clear that the transfer of classification accuracy between the calibration and the application phase needs to be improved. One way of accomplishing this is by using adaptive classifiers, which use new, incoming data to update the classification model (Shenoy, 2006; Galan, 2008). The disadvantage of this approach is that it requires immediate labelling of this data, which in turn requires knowledge of the monitored mental state through different information sources. One way to get this information would be to simply ask the user if workload classifications are accurate. However, adding a task to the primary task is usually not recommended by researchers from the adaptive automation literature (e.g., Scerbo, 1996). Alternatively, discrepancies between expected performance levels and classified workload could be used to update the model. For example, good lane keeping performance combined with a slow moving car and high workload classifications, should be an indication that these data segments have to be labelled anew. Another way of improving accuracy transfer between a calibration and an application phase may be to preselect signals that the classifiers can use, thereby excluding sources of non-stationarity. In other words, theory- and data-driven approaches could be combined to create an optimally reliable system.

As is clear, workload classifications often did not seem to reflect the actual workload as experienced by the drivers. Therefore, the system did not react as expected, which is reflected by user experience ratings. Several lessons can be learned for future research. From an ergonomic point of view, including performance based decision criteria (the 'hard' safety loop) is critical for preventing a physiology based loop from setting extreme system states (e.g., driving very fast), as long as physiology cannot be used to assess the mental state of the system's user reliably. Alternatively, monitoring workload could be used by different types of support. For example, it could inform the driver of his/her inferred workload or use these to formulate an advice to the driver. In this way, the system would support the earlier stages of the human information processing system (e.g., Parasuraman et al., 2000; Prinzell et al., 2002) but keep the ultimate decisions and vehicle control in the hands of the human users. Speculatively, drivers may sooner accept a somewhat less reliable system if it is not directly coupled to vehicle control.

In conclusion, while the classification models that were trained on different levels of visuomotor workload could accurately classify between three data classes of the calibration phase, results from the application phase indicated a low transfer of calibration accuracy. However, during the application phase, the vehicle parameter loop that monitored lane-keeping performance, prevented the system at large to increase speed to extreme speeds. This study confirms that substantial advances are required before workload can be reliably monitored in these conditions.

# Chapter

# 8

# Main Results and Discussion

Sensor technology and computing power have progressed to the stage that systems developed for inferring emotions, intentions, and mental capacities may be used to support our activities. As these systems evolve, the experience of interacting with this new generation of technology will increasingly resemble human-human interaction. The rationale behind these developments is that technology capable of recognising our mental state is better able to help us achieve goals by adapting to our changing wishes and needs (Brusilovsky, 2001; Brusilovlsky and Millán, 2007; Duric et al., 2002; Feigh et al., 2012; Hettinger, 2003; Fairclough, 2009). The research presented in this thesis aims at contributing to this vision of adaptive human-machine interaction by focussing on a specific mental state and task performance environment. To be precise, mental workload changes as a result of driving a car on roads in rural conditions and subsequently, how these inferences may be used by a driver assistance system in real-time to aid the driver. In this chapter, the most important findings from the studies presented in this thesis will be summarised and discussed separately, before providing a general discussion and formulating the main challenges for future research.

## 8.1 Chapter 3. Steering Demand and Mental Workload

The main objective of this study was to explore how various levels of lane keeping demand, which were created by manipulating lane width and oncoming traffic density on bendy, rural roads, affect mental workload as measured by steering behaviour, subjective effort ratings and heart rate indices. The findings of this study indicated differences between measures with respect to sensitivity. That is, how well a change in mental workload also results in a change of the measure, and with respect to the range of tasks demands for which measures are sensitive.

To start with, the lateral position (SDLP), indicating the level of engagement of the human visuomotor system in keeping the vehicle at the preferred lateral position on the road, turned out to be sensitive to all levels of lane keeping demand. SDLP decreased about 0.01 m per 0.25 m decrease in lane width from 3.00 m to 2.25 m. A stream of high oncoming traffic density (40 passing cars per minute vs. 10) further decreased SDLP by about 0.04 m. This effect on swerving behaviour can be seen as a behavioural adaptation by the participants in a (subconscious) attempt to maintain safety margins across driving situations (e.g., Summala, 2007). A low SDLP, especially combined with a shift of the car's lateral position towards the shoulder, which was also found, minimises the number of close proximities to oncoming traffic. Subsequently, a driver can feel comfortable again. According to mental workload theories (e.g., de Waard, 1996), maintaining safety margins in conditions of increasing task demand can only be realised at the cost of simultaneously increasing mental effort. This was also reflected by mental effort ratings and heart rate variability, but not in the same way.

Mental effort ratings increase reflected both increases in traffic density and decreases in lane width. However, the increase as a result of the lane width manipulation only occurred for the high-density traffic condition, which indicates that the sensitivity of experienced mental effort ratings is in a somewhat different range than SDLP. This divergence is in accordance with mental workload theory, which suggests that workload is hardly felt for intermediate levels of task demand. The effort expenditure that is required for task performance is only experienced outside this range (see also de Waard, 1996). An explanation for the effort rating findings is therefore that the increase in traffic density, which had clear effects on mental workload related driving behaviour, pushed task demand outside the optimum range, after which the additional lane keeping demand increases were also felt.

Finally, the 0.10 Hz component of heart rate variability, which has been used extensively in mental workload research (e.g., Mulder et al., 2009), differed as a result of different traffic densities. However, the expected power decrease in this frequency band (0.07 - 0.14 Hz) as a result of increased lane widths was not found, which indicates that this measure was the least sensitive to variations in mental workload. The observation that different measures differ in sensitivity has been seen in other driving contexts as well (e.g., de Waard, 1996) and therefore confirms that using multiple types of measures is probably required for reliable mental workload assessment over a wide range of task demands when designing a support system.

## 8.2 Chapter 4. A Performance Based Adaptive Driver Support System

Given that vehicle parameters turned out to be the most sensitive to changes in mental workload, it was decided to focus on SDLP and related parameters to design and test an interactive driver support system. This required that general knowledge at the group level, attained from the first empirical study, needed to be implemented at the individual level.

For this study, participants received information about the car's lateral position through a head-up display, but only when deemed necessary by the system. In terms of adaptive automation systems, this system supported the driver in the first stage of the human information processing system: acquiring information, and not in later stages such as making decisions or controlling the vehicle (e.g., Parasuraman, 2000). The roadway environment was very similar to the roadway environment of the previously discussed study. That is, lane-keeping demand was varied by manipulating oncoming traffic density (10 vs. 40 passing cars per minute), and driving lane width (2.25 m and 3.00 m wide). The system kept track of vehicle parameters over a 30 s moving window and activated the head-up display when SDLP was over 0.22 m, or when the participants drove either too near or outside the lane edges for more than 7.5 s or 3 s respectively.

The results indicated that driving with this user adaptive system slightly improved driving performance compared to driving without support and driving with a continuously activated (non-adaptive) support system. However, these effects were small compared to the lane width and oncoming traffic manipulations. Also, effort ratings did not differ between support type conditions. User experiences may provide a way to explain these results. Surprisingly, almost one third of the participants chose to ignore the projected information, presumably because they did not find it useful or satisfactory. This behaviour might be classified as disusing the provided assistance, which can be considered as a failure to engage in automation when it could improve performance (Parasuraman & Riley, 1997; Lee, 2008).

However, simply observing why such a large group chose to ignore the information does not explain why it happened. The effective implementation of visual aid may depend on many factors, for instance HMI, including the actual visual aspects (such as attractiveness, colouring, size, location on the windscreen, etc.), the exact triggers and trigger values (e.g., a strict vs. a tolerant system), and it may also take some time for a driver to learn how to use a system and to optimally acknowledge its advantages. A thorough investigation of all factors involved may provide options to improve the user experience substantially. However, it should be noted that, even in its current form, 16% of the participants did actually indicate to have used the information as intended. Considering the large proportion of the population that engages in car driving, this shows that there already is already a substantial group of drivers that may benefit from the setup in the present context.

Furthermore, participants indicated that the adaptive system was appreciated more than the non-adaptive version, because the moment of support activation was perceived as a warning signal that lateral control should be improved. This points to the fact that the system did not only support the information acquisition stage, as was the intention of the designers, but also a later stage of the human information processing: making a decision with respect to driving style. This should be kept in mind when designing a user adaptive system. Users of a system will probably think about how the system "thinks" and use this information to adapt their behaviour.

## 8.3 Chapter 5. The Potential of Music Selection for Adaptive Driver Support

In search of alternative ways of influencing driving behaviour, the feasibility of using negatively and positively rated music by an adaptive driver system was investigated. Two opposite effects of music listening while driving could be expected from the literature. On the one hand, an arousing effect in monotonous driving circumstances may counter a decreasing workload capacity due to under-arousal (e.g., North & Hargreaves, 2008). On the other hand, music may capture attention resources, which should be used for keeping the vehicle safely on the road, and have a negative impact on driving behaviour (e.g., Shek & Schubert, 2009). During this study, participants were therefore exposed to a relative high demanding drive in addition to a relative low demanding drive. Similar to the previously discussed chapters, these conditions were implemented by exposing each participant to a 2.50 m and a 3.00 m wide driving lane condition.

Since musical preference is highly personal, a particular piece of music may be perceived very differently across individuals. To compensate for this variability, participants first rated about 60 songs from a broad musical spectrum during an initial session. A compilation of negatively rated and a compilation of the positively rated songs were then created while attempting to match for rated energy levels. As it turned out, the selected songs indeed differed in valence, but they also differed to a lesser extent in energy levels. Within the range of songs used for this study, it was impossible to select songs that solely varied in mood valence showing that valence and energy ratings were positively correlated.

In the original paper of this study (Van der Zwaag et al., 2012) the central question was whether or not a specific mood could be successfully induced and maintained during driving through music listening, to which the answer was a modest yes. However, the effects of listening to music on driving behaviour, physiological measures, and mental effort ratings were limited to marginal effects on speed and respiration rate. Speed was somewhat lower during driving while listening to positively rated music compared to the no-music condition. In general, decreasing driving speed gives the driver more time to read signs, assess traffic situations, but also to make steering wheel corrections. An explanation for this result may therefore be that drivers listened intently to positively rated music, drawing attention away from the driving task, causing a (subconscious) compensatory speed reaction (e.g., Summala, 2007). Respiration rate also showed an effect of music as it slowed down when drivers listened to negatively rated music. This could indicate that listening to music may have had a relaxing effect on drivers.

In conclusion, it seems unlikely that automatic music valence selection by a driver support system could be used to directly change mental workload under the driving conditions of this study. However, the speed reduction due to listening to positively rated music is an indication that attention resources are captured by music, and therefore, switching off music in high demanding situations could be considered for future research. The influence of listening to music on respiration rate could indicate a more direct effect on bodily and affective states such as arousal level. This suggests that automatic selection of in-vehicle music could aid the driver in regulating arousal or relaxation levels, which could in turn affect driving safety. However, more research is required to confirm the effectiveness of these suggested adaptive strategies.

## 8.4 Chapter 6. Classifying visuomotor workload from brain waves

This study was, again, aimed at finding sensitive measures of mental workload on rural road conditions albeit through a very different data analysis approach than the approach used in the previous studies in this thesis.

Participants of this study completed fifteen short rides, with varying levels of steering demand. Steering demand was manipulated by changing driving speed (5 levels) and by providing the participant with performance target levels (3 levels), all relative to the individual's comfortable driving style as established during an initial ride. For all rides, traffic interaction was limited to a stream of 40 passing oncoming passenger cars per minute on average. Performance targets of swerving behaviour (SDLP) were projected onto the windscreen of the vehicle in addition to current values of SDLP, as measured during a 15 s moving window. Unexpectedly, the steering performance target manipulation did not affect the actual swerving behaviour as was the case with lane width and traffic density manipulations during the previous studies (see chapter 3 and 4). On the one hand, this indicates that the intended mental workload manipulation failed, since SDLP proved to be the most sensitive measure in similar driving circumstances. On the other hand, the most difficult performance target was rated as more effortful, suggesting that drivers were trying but unable to comply with the task instructions. It may be that such an explicit instruction is difficult to translate to activities of the largely automated visuomotor system. To the very least it shows that SDLP was not sensitive to this type of task demand, which again confirms that a single measure may not suffice to assess mental workload over a broad range of task demands.

While driving, the participant's EEG was recorded additionally for an offline workload classification study on the individual level instead of on the group level. The other main feature of this approach is that it is data-driven. That is, EEG data were transformed through the use of machine learning algorithms in such a way that the discriminability between experimental conditions was maximised.

The workload classification procedure used in this study involved four main steps, which were carried out for each participant individually. Firstly, the eye-movement corrected EEG data were segmented into one-second epochs. Secondly, the Common Spatial Pattern (CSP) technique created a unique linear combination of the EEG data from two selected experimental conditions. This maximises the difference in the signals' variance between one condition and the other (Blankertz et al., 2008). Thirdly, the variance values of each data-epoch were entered into a Fisher's Linear Discriminant Analysis (LDA), which again transforms the data in such a way that the distance between data points between conditions are maximised. These three first steps, of training the classifier, were carried out over a randomly selected, large portion of data epochs, leaving the remaining data for testing the classifier. Finally, these steps were repeated a large number of times in a cross validation procedure to avoid a data selection bias.

The results of this study showed that this technique is highly sensitive to changes in mental workload as classification accuracy increased to 95% on average across participants for the higher frequency bands when discriminating between experimental conditions. This was a promising finding, confirming that this data analysis method may be used to develop EEG-based adaptive support systems.

However, the superior classification performance resulting from high EEG frequencies raised the question of the neurophysiological mechanisms underlying these classifications. From the literature, it is known that EEG frequencies above 20 Hz are often contaminated by electrical activity resulting from muscular activity. The association between high accuracies in high frequency bands therefore suggests that part of the underlying information did not originate directly from neuronal activity. This holds especially in case of a semi-realistic task such as driving in a simulator, which requires a lot muscular activities compared to more strictly controlled laboratory tasks. Also when looking at scalp topographies, which reflect how the CSP-transformed signals project to the scalp, it is clear that peripheral electrodes, which are most likely to pick activities from facial muscles, were often involved in case of very high accuracies. Nonetheless, for EEG frequencies below 20 Hz, which are less likely to be contaminated by EMG activities (Whitham et al., 2007), an average accuracy of up to 80% was still found, indicating that a user adaptive support system that is largely based on neural activities may still be feasible (e.g. passive BCI).

Distinguishing between workload levels through the classification technique described above differs from more traditional statistical approaches, which hinders a clear comparison between workload measurement techniques. As mentioned, the CSP-based procedure transforms data at the individual level to maximise discriminability. In addition, it is fundamentally limited to classifying between two data classes (i.e. two workload levels). Nonetheless, such high accuracies do indicate that the EEG classification technique is more sensitive to a broader range of mental workload changes than either SDLP or subjective ratings.

## 8.5 Chapter 7. A Brain and Performance Based Adaptive Cruise Control

For this study, two versions of a proto-typical passive BCI system were implemented in a driving simulator, which were aimed at maintaining workload levels by increasing, maintaining, or decreasing driving speed. Strictly speaking, this system cannot be defined as adaptive automation, since it did not shift tasks back and forth between the human user and a technological subsystem (see Chapter 2 for more details on adaptive automation). However, it was certainly designed to be user adaptive and fits well within the general philosophy of adaptive automation to keep workload levels within an optimal range.

The reason for using two versions of a BCI was our interest in creating a BCI that is largely based on neuronal activities, as well as creating a BCI that simply performs best in terms of classification accuracy. In the literature there are clear indications that EEG frequencies above 20 Hz are contaminated with electrical activities as a result of muscular activities. However, electromyogram (EMG) activity is largely absent from EEG frequencies below 20 Hz (Whitham et al., 2007). Therefore, a low frequency BCI system version was tested in addition to a high frequency BCI version, while participants drove through rural scenery that varied with respect to curve length to manipulate steering demand.

Since this brain-based adaptive cruise control needed to distinguish between three levels of workload during the application phase, the binary workload classification procedure described in Chapter 6 needed to be expanded to a multiclass classification procedure. This was accomplished through a pairwise voting procedure (e.g. Friedman, 1996). That is, during the calibration phase, data from three workload levels were acquired (low, comfortable, and high), relative to the individuals' comfortable workload level as established during an initial ride. Next, three CSP-based classifiers were trained (per BCI version), one classifier for each condition pair. Subsequently, during the feedback phase, all three classifiers were deployed simultaneously and assigned probabilities to each new one-second EEG segment indicating the likely hood that data should be classified as A or B. In other words, all data classes (low, comfortable, and high workload) received two probabilities, which were then added up. The data class that received the highest summed probability "won" the vote, resulting in a speed change advice after applying a moving average for these probabilities to further enhance classification accuracy.

As a precautionary measure, vehicle parameters were also monitored during the application phase, using a 20 s moving window, allowing the speed regulation protocol to be expanded with vehicle-based interventions. In addition, an absolute minimum and maximum driving speed was set and a change dead-band (10 s) was implemented, which prevented very rapid speed changes and allowed the driver some time to get used to a new speed.

The calibration data were not only used for model training, they were also assessed offline after the experiment had finished and accuracy levels roughly replicated the accuracy as found during the preparatory study (Chapter 6). However, before discussing the results from the feedback phase, it should be noted that the workload classifications that were used by the adaptive system are now suspected to have had time delays for a number of participants. This delay could range from about 1 s to a maximum of about 5 s. Even if the actual impact of these delays may have been relatively low due to the dead-band that limited the system's responsiveness, an immediate link between workload classifications and driving speed cannot be assumed. This limits the conclusions that can be drawn from this study.

Having said that, upon applying the trained classification models to new EEG data in real time during the application phase, a wide variety of (sometimes erratic) speed setting behaviours were observed, indicating that classification accuracy had dropped substantially. For example, for several rides, the BCI system constantly advised the driving simulator to reduce speed. As a result, driving speed was maintained at the absolute minimum level. Another example is an oscillating speed pattern caused by workload classifications that continuously advised to increase speed but was subsequently overruled by the vehicle parameter loop, which signalled worsened lane keeping performance. However, workload classifications of comfortable levels and an oscillating pattern around this level were also observed, indicating that for some of the rides, the system behaved as expected. Systematic differences in system behaviour between the BCI version and curve lengths were not observed.

It must be concluded that the CSP-based classification model, as implemented for this study, turned out to have low transferability going from the calibration to the feedback phase, which severely impaired the technique's ability to reliably monitor mental workload. In general, the time and context dependency of psychophysiological measures is a known issue, both within the field of BCI and in the wider human-machine interaction field (e.g., Shenoy et al., 2006; Fairclough, 2009). In case of passive BCI it shows that substantial improvements of accuracy transfer, for example by updating the classification models in real-time while they are being applied, should continue to be one of the main research topics in this field.

## 8.6 General discussion

The main goal of this thesis was to find ways to monitor changes in mental workload levels as elicited from the lane-keeping task when driving on rural roads. So far in this chapter, the studies aiming at this general goal were discussed separately. Upon arriving at this point in this thesis however, I will address some of the more general issues that, in my opinion, may be relevant for a broader range of research into advanced human-machine interaction. This range of research is mainly defined by its goal: create applications that can act on what they infer from the user.

Monitoring mental workload requires knowledge of the associations between a construct that is defined in psychological terms and the sources of information that may be available to the workload monitor. It could be argued that there are two basic research paradigms through which this knowledge may be attained: the offline and the online paradigm. Firstly, in offline studies the effects of experimental manipulations are analysed when the data from all participants are recorded. Offline studies can be divided into studies using statistical approaches aimed at drawing conclusions at the group level and at the individual level. Secondly, there are online studies during which the on-going state-changes of the user are monitored while a participant is exposed to the study's conditions; and usually also a form of feedback is included. That is, a detected change in the user state can lead to a machine action, such as providing information, warnings, or taking control of the task (e.g., adaptive automation; see Chapter 2 for an overview). Of course, online monitoring can also be carried out without any form of interaction, but this type of study was not part of this thesis and will not be discussed any further.

In the human factors literature in general and also more specifically in the transport psychology literature, the offline paradigm at the group level is most commonly used. Examples of investigated mental states include fatigue, distraction, alertness, perceived risk, workload, emotions and so forth (Åkerstedt, 2004; Thomas & Walton, 2007; Brookhuis & de Waard, 2010; Lewis-Evans, 2012; Zhao et al., 2012; Strayer et al., 2013; and many more). Also the first, third, and fourth study of this thesis (Chapters 3, 5, and 6), fit into this paradigm. These experiments showed which of the included measures (subjective, performance, and physiological) reflected mental workload and in what way. A performance measure, the standard deviation of the lateral position, turned out to be highly sensitive to most task demand manipulations, but not all. These differences in the measures' sensitivity show there is no single measure that will reflect all levels of mental workload under all driving conditions, not even when the task is restricted to driving on rural roads in a simulator (de Waard, 1996; see also Figure 1.2). In other words, a many-to-one measures-state mapping is required to establish a reliable relationship (see Fairclough, 2009; Cacioppo and colleagues, 1990, 2000 for different types of measure-state mappings, albeit in the context of psychophysiological measures only).

So far, what these offline analyses have in common, is that they rely on traditional statistical inference, which is arguably not the best approach for uncovering relationships that can be directly used in online studies. In a nutshell, traditional statistical analyses usually compare the differences between experimental conditions (e.g. workload levels) to the differences within these conditions, to infer the likelihood of the found differences between conditions. The differences between individuals are thereby marked as error variance. Even in a repeated measures analysis of variance, where the variations due to individual differences are partly taken out of the error term, the directions of effects within the individuals need some consistency across individuals to reach statistical significance. In other words, traditional statistical approaches are designed to draw conclusions at the group level. When the research goal is to generalise to the larger population, for example to advise policy makers in sectors such as transportation, health care, and industry, group-level statistics may be the most appropriate approach. However, when we are mainly interested in the individual's emotional or cognitive state, the individual differences are exactly what we have to deal with.

One way of dealing with individual differences is through the feature extraction and classification algorithms that are commonly used in BCI research. The main advantage of these machine learning approaches is that they are data driven in the sense that data from one individual can be used to classify between mental states, such as mental workload levels, without any preconceptions of the fundamental (neurophysiological) mechanisms underlying the differences between data from one mental state and the other (e.g., Blankerz et al. 2008; Müller et al.; 2008; Zander et al., 2011). Apropos, that is not to say that assumptions of the underlying mechanisms are not implicitly or explicitly inserted by the researcher through the selection of input data channels or the selection of EEG-frequency bands.

Both in the ergonomic and BCI related literature, examples of various types of machine learning algorithms for offline analysis can be found, ranging from artificial neural networks to relative simple linear classifiers (e.g. de Waard et al., 2001; Wilson and Russel, 2003; Müller et al., 2008; Hockey et al, 2009; Wang et al, 2012). The EEG analysis described in Chapter 6 also fit into this research approach, as the CSP technique was applied to extract the most relevant features before applying Fisher's linear discriminant analysis. This approach was highly sensitive to all levels of mental workload elicited during the experiment. Given these outcomes and the individual nature of the technique, this approach seemed very promising to investigate during an online study.

For online studies, the main research goal is to determine the condition of the individual on the fly, which poses several challenges to the researcher compared to offline studies on the group level. Firstly, similar to offline classification studies, the relational direction is reversed. It is now necessary to assess the level of the independent variables through the response pattern of the dependent variables. Secondly, an online measurement technique requires more robustness. The usefulness of assessments does not just depend anymore on the existence of a reliable and sensitive association between the two types of variables, it is also required that the measure is selective. Selectivity of a measure refers to the degree that a measure only changes if the mental state changes, and is not dependent on other factors (de Waard, 1996; see also Fairclough, 2009).

Thirdly, monitoring needs to be carried out at the individual level and therefore knowledge that was gained from group-level analysis needs to be translated to online mental state inferences. In the literature, several methods to accomplish this can be identified. To start with, a distinction can be made between absolute and relative criteria (Brookhuis et al., 2003). Absolute criteria are identical for all individuals, but are nonetheless applied at the individual level. For example, Brookhuis et al. (2003) published a paper providing absolute criteria for driver impairment (in addition to relative criteria), which could be used to trigger a machine action. Also in this thesis (Chapter 4 and 7), absolute lane keeping performance criteria were used to time the provision of driving support, which are signs of inadequate mental effort expenditure leading to unsafe driving behaviour. Relative criteria refer to baselined measures. In this case, mental state inferences are based on the measure's differences between a control situation and the current situation. For example, the bio-cybernetic system used by Prinzel et al., (2001) compared the values of the so-called 'engagement index' while performing a task, which is a ratio between several EEG frequency powers (beta / (alpha + theta)), to baseline (i.e. resting) values of the same index. Slightly different, in the original engagement index study, the decision to provide or withdraw automation was based on the slope of the index, which was updated every couple of seconds (Pope et al., 1995).

The various methods of criteria setting described above are still characterised by very general threshold rules, identically applied to all users. Machine learning algorithms on the other hand, are able to transform input channels in such a way as to maximise the differences between mental states at the individual level. In essence, a new set of criteria is derived for every individual. Several examples of online systems that were geared up with a machine learning approach can be found in the literature (Kohlmorgen et al., 2007; Wilson & Russell, 2009; Ting et al., 2010; Christensen & Estepp, 2013), although their number is limited compared to the number of offline classification studies. In general, online classification studies are characterised by a training phase and a feedback phase. During the training or calibration phase, data are acquired while a mental state is manipulated through variations in task conditions, similar to offline studies. In other words, input data are labelled as reflecting a particular state, such as low, comfortable, and high mental workload. Then, the parameters of the classification model are determined, based on the data from two of these task conditions (e.g., weights are added to the original EEG channels to maximise discriminability between high and low workload). During the subsequent feedback or application phase, the same classification model parameters are again used to transform new, incoming segments of EEG data into mental state classifications.

This approach was also used for the online study described in Chapter 7, when online workload classifications were linked to a cruise control. The results from this study showed that very high offline classification accuracy as found during online studies does not mean that the same accuracy is transferred to the online paradigm. This is a common finding, and often attributed to non-stationarities in the EEG signal (Shenoy et al., 2006; Lotte et al., 2007; Krauledat, 2008; Borghini et al., In Press). Another way of framing this problem is to say that the machine learning algorithms that enable us to construct the much needed individually trained classification models, are also trained by differences in the EEG signal that are caused by factors that are unrelated to the relevant mental state, such as time-of-day differences. As mentioned before, tackling this problem, for example by using adaptive classifiers that are updated during a feedback phase, is required to develop reliable state monitors. To accomplish this, further cooperation and integration between the fields of Physiology, Human Factors and Ergonomics, and Computer Science is strongly needed.

When designing an adaptive system for an online study, one is quickly faced with the necessity to implement precautionary measures to deal with unlikely events. Even if online workload classification accuracy could be increased to 95%, there would still be 5% misclassifications, which could have detrimental effects on task performance, user acceptance, and in case of driving, on traffic safety. For example, complacency is often mentioned in the literature to indicate that a human user will tend to over rely on a technological system that is not completely reliable, and will therefore be more likely not to notice when the system is at fault. Similarly, automation bias may lead human operators into making decisions that are strongly influenced by the technological system's advice for example, instead of thinking critically for themselves (e.g., Parasuraman & Manzey, 2010). Needless to say, these human tendencies will have to be factored in when designing a support system that is linked to a user monitor that has some degree of unreliability, as is the current state of affairs. One way to deal with this problem is to combine physiological data with other data sources such as task performance to prevent extreme system states (e.g., very fast driving speed). In general, the added value of this type of user adaptive systems will likely depend on the combination of monitoring reliability and the specific application which the systems are connected to.

So far, this general discussion has only dealt with issues with regard to monitoring the mental state of the human user. However, as is clear from the adaptive automation literature, this is just one part of the equation towards user adaptive systems. The other part entails issues with regard to the machine actions. Especially, the way that the user reacts to an adaptive system is a necessary research direction. Behavioural adaptions from the human user can be expected as users gain experience with the system, or through misuse, disuse, or abuse of the system, facilitated by human capacity to adapt to new situations (e.g., Parasuraman & Riley, 1997; Hancock & Verwey (1997). However, developing reliable state monitors for the online situation should be seen as prerequisite before investigations into behavioural adaptations to these systems make sense.

Now, finally, upon nearing the end of this thesis, what else may be said? In the introduction, it was stated that this thesis is dedicated to the advancement of human-machine interaction. How can we reliably monitor human mental state and how can monitoring aid the human user? It is clear that we can use many sources of information and many data analysis techniques to monitor the mental state, and none of them are perfect. However, by investigating a number of these possibilities, such as several driving behaviours, physiological measures such as heart, breathing, and brain activities, by analysing these on the individual level and using these insights to create two user adaptive systems, this thesis should be seen as another step forward towards user adaptive systems. Based on the research in this thesis, the next step would be focussed on increasing the monitor's accuracy over time and between situations. This may be accomplished by further exploring data-driven techniques. However, at the same time we should continue to investigate which input data-channels are in fact informative of the mental state that we are interested in, to be able to pre-select information sources for a mental state monitor, since it may very well be that a purely data-driven approach that combines as much measures as possible will turn out to be too specific to be helpful. I hope that these considerations may be beneficial in formulating future research efforts. If so, this thesis will have truly accomplished its goal.

# Nederlandse Samenvatting

# Dutch Summary

## Hoofdstuk 1. Introductie

Het centrale thema in dit proefschrift is gebruikers-adaptieve mens-machine interactie. Dit type systemen wordt ontwikkeld vanuit de visie dat in de nabije toekomst technologie in staat zal zijn onze emoties, intenties en mentale capaciteiten af te lezen en dat de omgeving daardoor flexibel aan te passen is aan onze voortdurend veranderende wensen en behoeften (Brusilovsky, 2001; Brusilovlsky and Millán, 2007; Duric et al., 2002; Feigh et al., 2012; Hettinger, 2003; Fairclough, 2009). In figuur 1.1 is dit idee in de meest eenvoudige conceptualisatie weergegeven, als een soort regelsysteem, verglijkbaar met een centrale verwarming waarbij een thermometer de verwarmingsinstallatie doet aan- en afslaan. Op een vergelijkbare manier gebruikt een gebruikers-adaptief systeem informatie om te bepalen of de gebruiker bijvoorbeeld overbelast is en daarom behoefte heeft aan ondersteuning. Deze optionele ondersteuning zorgt er vervolgens weer voor dat de belasting voor een gebruiker op een aanvaardbaar niveau komt, enzovoort. Een toepassing van dit concept is een auto met een vermoeidheidsdetector, die de bestuurder suggereert een pauze te nemen.

Het onderzoek dat in dit proefschrift beschreven is, werd opgezet met als doel bij te dragen aan deze visie van adaptieve mens-machine interactie. Deze bijdrage werd geleverd door de nadruk te leggen op een bepaalde mentale toestandsverandering in een specifieke taakomgeving. In het bijzonder ging het bij dit onderzoek om het detecteren van mentale inspanningsveranderingen als gevolg van autorijden (in een rijsimulator) op verschillende wegen. De vervolgvraag was hoe informatie over deze inspanningsveranderingen gebruikt kan worden om de bestuurder te ondersteunen terwijl hij aan het autorijden is.

## Hoofdstuk 2. Theoretische achtergronden

Het algemene idee van gebruiker-adaptieve technologie vond zijn oorsprong in de jaren zeventig van de vorige eeuw, zoals beschreven wordt in de vakliteratuur van een subcategorie van het ergonomisch onderzoeksveld: de adaptieve automatisering. Deze tijd werd gekenmerkt door de opkomst van de automatisering, die vele voordelen bood ten opzichte van de handmatige taakuitvoering. Echter, men constateerde ook een aantal negatieve effecten die gerelateerd waren aan de nieuwe rol van de mens in dit geautomatiseerde systeem. Volautomatische systemen moeten gemonitord worden door een menselijke supervisor voor het geval er noodsituaties ontstaan. Kortgezegd, de rol van de mens veranderde van een actieve taakuitvoerder naar een passieve observant, die slecht uitgerust is om de sporadisch optredende fouten in het automatische systeem op te vangen, bijvoorbeeld doordat de eentonigheid van deze taak de alertheid vermindert (Sheridan, 1976a, 1976b; Sheridan & Verplank, 1978; Wiener & Curry, 1980; Bainbridge, 1983; Wiener, 1989; Endsley, 1995; Billings,1997; de Waard et al., 1999). Dit nadeel leidde tot het idee dat het beter is om automatisering flexibel in te zetten. Namelijk, op een zodanige manier dat de menselijke uitvoerder alleen ondersteund wordt indien hiertoe voldoende aanleiding is, waardoor de betrokkenheid bij het primaire taakproces zo groot mogelijk blijft.

In de literatuur over adaptieve automatisering staan twee onderzoeksvragen centraal. Ten eerste: wanneer moet een adaptief systeem iets doen? En ten tweede: wat kan een adaptief system doen? Anders gezegd, wat moet aangepast worden en wanneer?

Wat aangepast kan worden wordt met name bepaald door de specifieke taaksituatie. Er zijn twee veelgebruikte theorieën die een ontwerper kunnen helpen hier over na te denken. Om te beginnen zijn er meerdere automatiseringsniveaus denkbaar waartussen geschakeld kan worden, waardoor de taaklast van de uitvoerder dynamisch kan worden aangepast. Dat wil zeggen, een automatisch systeem hoeft niet perse volautomatisch te zijn. Ten tweede, taakuitvoering wordt veelal gekarakteriseerd door een aantal fasen, zoals informatie-acquisitie, informatie verwerking, beslissen dat een bepaalde handeling nodig is en als laatste, de uitvoering van een handeling. De automatisering of taakondersteuning van elk van deze fasen kan ook flexibel van aard zijn. Een adaptief systeem kan bijvoorbeeld worden ingezet om op het ene moment alleen de handeling uit te voeren waartoe de mens (operator) besloten heeft, maar op het andere moment kan het systeem ook handelingssuggesties geven (Endsley & Kaber, 1999; Parasuraman et al., 2000; Sheridan, 2000).

Deze veranderingen van automatiseringsniveaus zouden door de menselijke taakuitvoerder geïnitieerd kunnen worden. Het gevolg hiervan is echter dat deze een extra taak krijgt, namelijk het aansturen van het flexibele ondersteuningssysteem. Het alternatief is dat het systeem zelf besluit wanneer het tijd is actie te ondernemen. Dit laatste type staat centraal in dit proefschrift. Voor een goede timing kan een systeem informatie gebruiken uit drie soorten bronnen. Ten eerste, informatie uit taakomgeving kan aangeven wat bijvoorbeeld iemands taaklast is (bijv. verkeersdrukte) of dat er zich een kritieke gebeurtenis heeft voorgedaan waardoor extra taakondersteuning zinvol is (bijv. een file verderop). Ten tweede kan in sommige taakomgevingen het menselijk gedrag een indicatie zijn van de taakprestatie. Gedragsobservatie kan ook bepalen of de handelingen die iemand uitvoert naar verwachting zijn. Indien de prestatie beneden de maat is of iemands handelen aangeeft dat de uitvoering suboptimaal verloopt, kan de computer ondersteuning aanbieden. Ten derde, kennis van de interne toestand van de mens, zoals stress of het mentale inspanningsniveau, kan een indicatie geven dat de taakuitvoering suboptimaal dreigt te worden, wat de aanleiding kan zijn voor het adaptieve systeem om een preventieve actie te initiëren. Kennis van de interne toestand kan onder meer verkregen worden door fysiologische metingen uit te voeren met behulp van sensoren op het lichaam (voor overzichten zie: Parasuraman et al. 1992; Kaber & Riley, 1999; Hoogeboom & Mulder, 2004).

### Hoofdstuk 3. Sturen en Mentale Inspanning

Deze studie was opgezet om te onderzoeken hoe rijmoeilijkheid de mentale inspanning beïnvloedt op bochtige, rurale wegen. Mentale inspanning werd afgeleid aan de hand van het stuurgedrag, subjectieve oordelen van de bestuurders en de fysiologie. Dit geeft, met andere woorden, verschillende maten voor mentale inspanning. Rijmoeilijkheid werd gemanipuleerd door de bestuurders bloot te stellen aan verschillende wegbreedten en door de dichtheid van het tegemoetkomend verkeer te variëren. De bevindingen van deze studie tonen aan dat de gevoeligheid voor veranderingen van mentale inspanning niet gelijk zijn voor alle gebruikte maten.

Allereerst bleek dat de standaard deviatie van de laterale positie die het voertuig innam op de weg (SDLP; het slingergedrag op de weg), gevoelig te zijn voor alle niveaus van rijmoeilijkheid. SDLP verminderde met ongeveer 0.01 m per 0.25 m versmalling in rijstrookbreedte (van 3.00 m tot en met 2.25 m). Een hoge dichtheid van de tegemoetkomende verkeerstroom op de andere weghelft (40 passerende auto's per minuut in plaats van 10), resulteerde in een extra afname/vermindering van SDLP van 0.04 m. Dit effect op het slingergedrag van de bestuurder kan gezien worden als een (onbewuste) poging om de veiligheidsmarges in alle rijsituaties gelijk te houden (Summala, 2007). Een lage SDLP, met name in combinatie met een verschuiving van de gemiddelde laterale positie op de weg richting de berm (wat ook gevonden werd), minimaliseert de frequentie dat de auto van de bestuurder dicht bij het tegemoetkomend verkeer komt. Hierdoor voelt de bestuurder zich weer op zijn of haar gemak. Volgens theorieën over mentale inspanning (zoals De Waard, 1996), is het handhaven van veiligheidsmarges (met andere woorden, de taakprestatie), gegeven verzwaarde taaklasten, alleen mogelijk indien de mentale inspanning evenredig toeneemt. Deze toename van mentale inspanning werd ook gezien in de subjectieve oordelen en in de hartslagvariabiliteit, maar niet op dezelfde manier.

De bestuurders scoorden hoger op de subjectieve mentale inspanningsschaal bij een toename van de verkeersdichtheid en daarnaast bij een afname van de rijstrookbreedte. De gevoeligheid van deze subjectieve maat voor de strookbreedte manipulaties beperkte zich echter tot de hoge verkeersdichtheid condities, wat aangeeft dat de gevoeligheid van deze maat in een ander bereik zit dan SDLP. Dit afwijkende patroon is in overeenstemming met theorieën van mentale inspanning, waarin gesuggereerd wordt dat mentale belasting (of inspanningseis) nauwelijks als inspannend wordt ervaren indien de taaklast middelgroot is (niet te licht en niet te zwaar). Inspanning wordt pas als zodanig ervaren wanneer de taaklast buiten dit optimale bereik komt (De Waard, 1996). Een verklaring voor het gevonden effect op de subjectieve scores is dus dat de toename van verkeersdichtheid, wat duidelijke effecten had op de SDLP, de taaklast verhoogde tot voorbij het optimale taaklastbereik, waardoor de additionele eisen aan de stuurprestatie ook bewust werden ervaren.

Verder, de 0.10 Hz component van de hartslagvariabiliteit, een veel gebruikte maat in onderzoek naar mentale inspanning (bijvoorbeeld Mulder, et al., 2009), reageerde op de toename van verkeersdichtheid. Echter, de verwachte afname van energie in deze frequentieband (0.07 – 0.14 Hz) als gevolg van minder brede rijstroken werd niet gevonden. Dit is een indicatie dat deze maat het minst gevoelig was voor deze veranderingen van mentale inspanning. De observatie dat inspanningsmaten verschillen in gevoeligheid is ook gevonden voor andere rijomstandigheden. Deze studie bevestigt hiermee het beeld dat er voor een gebruikers-adaptief systeem waarschijnlijk meerdere maten nodig zijn om een betrouwbare en gevoelige meting van mentale inspanning in een breed bereik van taakeisen te bewerkstelligen.

### Hoofdstuk 4. Adaptieve rijondersteuning op basis van rijprestatie

Uit de vorige studie bleek dat voertuigparameters het meest gevoelig zijn voor mentale inspanningsveranderingen. Daarom werd besloten om SDLP en andere, daaraan gerelateerde voertuigparameters te gebruiken voor het ontwerp van een rijondersteuningssysteem. Dit betekende dat de algemene kennis (op groepsniveau) verkregen uit de eerste empirische studie, nu toegepast diende te worden op de individuele gebruiker.

De rijondersteuning bestond uit informatievoorziening. Deelnemers aan deze studie kregen informatie te zien over de laterale positie van het voertuig op de weg, maar alleen indien het automatische systeem dat nodig achtte. De informatie werd geprojecteerd op de virtuele voorruit in de vorm van een ‘head-up display’. In termen van adaptieve automatisering kan gezegd worden dat dit systeem de eerste fase van het menselijk informatieverwerkingssysteem ondersteunde: informatie-acquisitie, en niet de latere fases zoals het nemen van beslissingen of het controleren van het voertuig (Parasuraman, 2000). De virtuele wegomgeving was in grote lijnen gelijk aan die van de vorige studie. De stuur-moeilijkheid werd gemanipuleerd door de dichtheid van de verkeersstroom uit tegenovergestelde richting te variëren (10 en 40 passerende auto’s per minuut) en door de rijstrookbreedte te veranderen (2.25 m en 3.00 m). Om tot de beslissing te komen of ondersteuning nodig was, gebruikte het ondersteuningssysteem alleen de meest recente halve minuut aan voertuigdata, dat gesampled werd met 10 Hz. Met elke nieuwe sample, schoof deze periode 0.1 s op in de tijd: een 'lopende' tijdsperiode, die telkens 29.9 seconden overlap had met de vorige periode. De informatievoorziening werd geactiveerd indien SDLP groter was dan 0.22 m, indien een deelnemer in totaal meer dan 7.5 s dichtbij de rijstrookgrenzen reed, of indien een deelnemer in totaal meer dan 3 s met (een deel van) het voertuig over de rand van de rijstrook reed.

Het bleek dat rijprestatie enigszins verbeterde met dit gebruikers-adaptieve systeem, in vergelijking met rijden zonder ondersteuning en rijden met continue informatievoorziening (non-adaptieve ondersteuning). Dit effect was echter klein vergeleken met de effecten van de wegbreedte- en verkeersstroomdichtheidsmanipulaties. De subjectieve mentale inspanningsoordelen lieten geen verschil zien tussen de drie ondersteuningscondities. De gebruikservaringen bieden een mogelijke verklaring voor deze resultaten want verrassend genoeg bleek dat bijna één-derde van de deelnemers de geprojecteerde informatie zoveel mogelijk had genegeerd, omdat deze niet als bruikbaar of bevredigend werd ervaren.

Alleen constateren dat een groot deel van de deelnemers er voor koos de informatie te negeren, geeft echter nog niet aan waarom zij dit deden. Effectieve implementatie van visuele ondersteuning hangt van allerlei factoren af, zoals de visuele aspecten van de interface (waargenomen aantrekkelijkheid, kleurgebruik, grootte, locatie op de voorruit, enzovoort). Ook hangt de ingeschatte bruikbaarheid van de ondersteuning af van de precieze drempelwaarden die bepalen wanneer ondersteuning geactiveerd wordt. Door deze aan te passen kan bijvoorbeeld een 'streng' of een 'tolerant' systeem gemaakt worden. Bovendien kan het zijn dat bestuurders simpelweg meer tijd nodig hadden om bekend te worden met het systeem en dus ook voordat de voordelen hiervan zouden worden ingezien. Een grondig onderzoek naar al deze factoren zou aanknopingspunten kunnen geven voor de verbetering van gebruikerservaringen. Tegelijkertijd mag niet onopgemerkt blijven dat 16% van de deelnemers aangaven dat zij het huidige, onderzochte systeem wel gebruikt hebben als bron van informatie. Omdat een groot deel van de bevolking wel eens auto rijdt, betekent dit dat een substantiële groep baat zou kunnen hebben bij een systeem van de huidige vorm.

Verder bleek dat het adaptieve systeem meer gewaardeerd werd dan de andere twee, niet-adaptieve, ondersteuningscondities, omdat het moment van ondersteuningsactivatie werd ervaren als een waarschuwingssignaal, namelijk, dat men zich meer diende te concentreren op de rijtaak. Dit geeft aan dat het systeem niet alleen de informatie-acquisitie fase ondersteunde, zoals de bedoeling was van de ontwerpers, maar ook een latere fase van het menselijk informatieverwerkingssysteem. Deze waarschuwing hielp de bestuurder namelijk met de beslissing wanneer zij voor hun gevoel hun rijstijl moesten aanpassen. Een technologisch systeem anders gebruiken dan de bedoeling was wordt in de literatuur onder de noemer van gedragsadaptaties geschaard. Gebruikers van een adaptief systeem zullen nadenken over instellingen van het systeem en deze informatie vervolgens gebruiken om hun eigen gedrag aan te passen. Immers, de mens is zelf ook een (biologisch) adaptief systeem.

### Hoofdstuk 5. De potentie van automatische muziekselectie voor adaptieve ondersteuning

Zoekend naar een alternatieve manier om rijgedrag te beïnvloeden, werd onderzocht of een gebruikers-adaptief systeem wellicht muziek zou kunnen gebruiken. Hiervoor werden twee typen muziek gebruikt. Muziek die een positieve- en muziek die een negatieve stemming bij de desbetreffende individuele bestuurder opriep. Op basis van de literatuur konden twee tegenovergestelde effecten van muziek luisteren tijdens het autorijden verwacht worden. Aan de ene kant zou het stimulerende effect een middel kunnen zijn tegen de afname van de mentale inspanningscapaciteit als gevolg van onder-activatie in monotone rijcondities, en dus een positief effect op rijgedrag kunnen hebben (North & Hargreaves, 2008). Aan de andere kant kon verwacht worden dat muziek aandacht vraagt die eigenlijk gebruikt moet worden om het voertuig veilig op de weg te houden; een negatief effect (Shek & Schubert, 2009). Daarom werden deelnemers aan deze studie blootgesteld aan zowel weinig- als veeleisende rijomstandigheden. Deze omstandigheden werden weer gecreëerd door wegstrookbreedte te manipuleren in de rijsimulator (2.50 m en 3.00 m).

Muziekvoorkeur is erg persoonlijk. Ieder individu zal een bepaald muzieknummer anders ervaren. Ter compensatie van deze variabiliteit, beoordeelden de deelnemers 60 nummers, tijdens een initiële sessie en werd geselecteerd uit een breed muzikaal spectrum, op de ervaren ‘stemmingsgevoeligheid’ van elk nummer (positief of negatief en in welke mate) en op het ervaren energieniveau. Vervolgens werd een compilatie van tien negatief beoordeelde nummers en een compilatie van tien positief beoordeelde nummers gemaakt, waarbij het energieniveau zo constant mogelijk werd gehouden. Achteraf bleek dat de compilaties inderdaad verschilden in stemmingsgevoeligheid, maar ook, zij het in mindere mate, in energieniveau. Het bleek niet mogelijk om nummers voor de compilatie te selecteren die alleen varieerden in stemmingsgevoeligheid, wat aantoont dat deze twee dimensies van emotie en stemming samenhangen.

Voor het oorspronkelijke artikel van deze studie (Van der Zwaag et al., 2012), was de centrale vraag of een stemming geïnduceerd en gehandhaafd kon worden tijdens het autorijden, door middel van muziek, waarop een voorzichtige bevestiging kon worden gegeven. De effecten van het luisteren naar muziek op rijgedrag, fysiologische maten en subjectieve mentale inspanningsmaten beperkten zich echter tot marginale effecten op rijsnelheid en ademhalingsfrequentie. In vergelijking met de condities waarin de deelnemers niet naar muziek luisterden, lag de gemiddelde rijsnelheid tijdens het beluisteren van positief beoordeelde muziek iets lager. In het algemeen blijkt dat bestuurders bij veeleisende weg- of verkeersomstandigheden vaak iets langzamer gaan rijden om zichzelf meer tijd te geven om bijvoorbeeld verkeersborden te lezen, een verkeerssituatie te overzien, maar ook om meer stuurcorrecties te kunnen maken. Een verklaring van het gevonden resultaat is dan ook dat de deelnemers aandachtig naar de positieve muziek luisterden, waardoor zij minder aandacht besteedden aan de rijtaak. Dit heeft mogelijk geleid tot een (onbewuste) compensatoire rijsnelheidsverandering (Summala, 2007). De ademhalingsfrequentie nam af (men ging langzamer ademen), wanneer men luisterde naar negatief beoordeelde muziek. Dit zou kunnen duiden op een ontspanningseffect.

Samenvattend lijkt het onwaarschijnlijk dat een automatische selectie van positief of negatief beoordeelde muziek door een adaptief systeem gebruikt kan worden om direct de mentale inspanning te veranderen in de rijomstandigheden die hier getest werden. De snelheidsverlaging bij positieve muziek duidt er echter wel op dat aandacht kan worden opgeëist. Een adaptief systeem dat automatisch de muziek uitschakelt wanneer er sprake is van veeleisende omstandigheden zou daarom een interessant onderwerp voor toekomstig onderzoek kunnen zijn. De invloed van muziek luisteren op ademhaling suggereert een direct effect op de lichamelijke toestand en bijvoorbeeld het alertheidsniveau. Muziekselectie in de auto zou de bestuurder daarom kunnen helpen bij het reguleren van zijn alertheid en ontspanning. De effectiviteit van deze mogelijke adaptieve strategieën moet echter nog wel verder uitgezocht worden.

### Hoofdstuk 6. Het classificeren van visueel-motorische inspanning uit hersengolven

Met deze studie werd weer teruggekeerd naar de primaire lijn in dit proefschrift: het vinden van gevoelige maten voor mentale inspanning tijdens het rijden op rurale wegen. Een groot verschil met de vorige studies is dat in dit onderzoek gebruik is gemaakt van een andere data-analyse benadering. Traditionele vormen van data-analyse worden altijd worden uitgevoerd op groepsniveau. Een vereiste van een gebruikers-adaptief systeem is echter dat deze op individueel niveau kan bepalen wanneer iemand behoefte heeft aan ondersteuning. Met andere woorden, een mentale toestandsmonitor kan geen gebruik maken van traditionele statistische (groeps-)methoden, terwijl de data-analyse methode die voor deze studie gebruikt werd wel op individueel niveau inzetbaar is.

Deelnemers aan dit onderzoek reden vijftien korte ritjes, die varieerden in stuur-moeilijkheid. Stuur-moeilijkheid werd gemanipuleerd door de rijsnelheid, die door de simulator bepaald werd, te variëren (op 5 niveaus) en door de deelnemers een prestatiedoel op te leggen (3 niveaus van moeilijkheid). De specifieke rijsnelheden en geëiste rijprestaties werden individueel ingesteld, op basis van de individuele, comfortabele rijstijl die tijdens een initiële rit bepaald werd. De interactie met ander verkeer was wederom beperkt tot een stroom van tegemoetkomende auto's (40 passerende auto's per minuut). Prestatiedoelen, uitgedrukt in een SDLP waarde (bijv. 25 cm), werden geprojecteerd op de voorruit, samen met de meest recente SDLP waarde die wederom berekend werd door middel van een 'lopende' tijdsperiode (vijftien seconden en tien maal per seconde opnieuw berekend).

Tegen de verwachting in bleek dat deze prestatiedoelen voor slingergedrag geen effect hadden op het eigenlijke slingergedrag. Aan de ene kant geeft dit aan dat deze manipulatie niet werkte, omdat SDLP juist heel gevoelig bleek voor andere manipulaties tijdens de andere studies. Aan de andere kant lieten de mentale inspanningsoordelen zien dat de bestuurders het zwaarste prestatiedoel wel degelijk als meer inspannend ervoeren dan de twee lichtere niveaus. Dit suggereert dat de deelnemers het wel probeerden, maar niet in staat waren de doelen te halen. Wellicht is zo'n expliciet prestatiedoel (getallen op de voorruit) moeilijk te vertalen naar de activiteiten van het grotendeels geautomatiseerde visueel-motorische systeem. Dit resultaat toont in ieder geval aan dat SDLP niet gevoelig was voor deze vorm van taakeisen, wat weer bevestigt dat één enkele maat meestal niet voldoende informatie geeft om de mentale inspanning te bepalen.

Tijdens het rijden in de simulator werd bij de deelnemers elektro-encefalogram (EEG) activiteit opgenomen. Het EEG is de grafische weergave van elektrische activiteit die gemeten wordt op de hoofdhuid. Deze wordt veroorzaakt door de onderliggende hersenactiviteit, maar ook door bijvoorbeeld oogbewegingen en het aanspannen van de gezichtsspieren. Deze data werden na afloop van het experiment gebruikt voor een classificatieonderzoek op individueel niveau. Tot dusver werden analyses juist gedaan op groepsniveau, waarmee de vraag werd beantwoord of mentale inspanningsmaten gemiddeld genomen onderscheid konden maken tussen verschillende inspanningsniveaus. Een andere eigenschap van de gebruikte analysemethode is dat deze data-gedreven in plaats van theorie-gedreven is. Dit betekent dat de EEG data van deelnemers zelf werden gebruikt als basis om filters in te stellen, bijvoorbeeld door sommige EEG kanalen (21 in totaal), zwaarder te laten wegen dan anderen. Deze instellingen (lees: het classificatiemodel), werden bepaald door wiskundige algoritmen waardoor verschillen in EEG data tussen experimentele condities gemaximaliseerd kon worden. Hierdoor werden de oorspronkelijke verschillen uitvergroot en kon makkelijker onderscheid gemaakt worden tussen bijvoorbeeld weinig of veel mentale inspanning.

De classificatieprocedure in dit onderzoek bestond uit vier stappen, die voor elk individu zijn uitgevoerd. Eerst werden de voor oogbewegingen gecorrigeerde EEG data gesegmenteerd in stukjes van één seconde. Daarna creëerde de Common Spatial Pattern (CSP) techniek een unieke lineaire combinatie van EEG data uit twee geselecteerde experimentele condities. Deze techniek maximaliseert het verschil in variantie van de EEG signalen tussen de ene conditie en de andere (Blankertz et al., 2008). Vervolgens werden de variantiewaarden van elk datasegment gebruikt voor Fisher's Lineaire Discriminanten Analysetechniek (LDA), waardoor de data opnieuw getransformeerd werden om verschillen tussen condities te maximaliseren. Dit eerste deel van de aanpak (eerste drie stappen), oftewel het trainen van het classificatiemodel, werd uitgevoerd op een willekeurig geselecteerd deel van de datasegmenten (75%). De overige segmenten werden gebruikt om de accuratesse van het model te testen ('weet' het classificatiemodel of een bepaald EEG segment uit de ene of de andere conditie komt?). De vierde en laatste stap was de cross-validatie procedure. Dat wil zeggen, de eerste drie stappen werden een groot aantal keren herhaald om te voorkomen dat de willekeurige selectie van de segmenten had kunnen leiden tot toevallige, niet representatieve testuitkomsten.

De resultaten van het classificatieonderzoek tonen aan dat deze techniek zeer gevoelig was voor veranderingen in mentale inspanning. Dit was af te lezen aan de stijging van de classificatieaccuratesse tot gemiddeld 95%, wanneer gekeken werd naar hogere frequenties in het EEG. Dit was een veelbelovende uitkomst van het onderzoek en bevestigt dat dit type data-analyse methode gebruikt zou kunnen worden door gebruikers-adaptieve systemen.

Deze superieure classificatieprestatie van met name de hoge EEG frequenties deed ook de vraag rijzen naar de neurofysiologische mechanismen die hieraan ten grondslag liggen. Uit de literatuur is bekend dat EEG frequenties boven 20 Hz veelal 'besmet' zijn met elektrische activiteit vanuit de spieren. De hoge nauwkeurigheid van het classificatiemodel in de hoge frequentiebanden suggereert daarom dat een deel van de onderliggende informatie niet direct afkomstig was van neuronale activiteit. Dit beeld wordt versterkt door de semi-realistische testomgeving van deze studie. Het besturen van de simulator vereist relatief veel spieractiviteit in vergelijking met de meer gecontroleerde laboratoriumopzet die normaal gesproken gebruikt worden voor EEG onderzoek.

Ook de schedelverdelingen, die laten zien hoe de getransformeerde EEG signalen geprojecteerd worden naar de schedel, laten zien dat de perifere EEG kanalen, die het meeste last hebben van activiteiten van de gezichtsspieren, vaak in sterke mate betrokken waren bij heel goede classificatieprestaties. Desalniettemin, ook voor EEG frequenties beneden de 20

Hz grens, welke in veel mindere mate beïnvloed worden door spieractiviteiten (Whitham et al., 2007), werd een gemiddelde accuratesse gevonden van tegen de 80%. Dit geeft aan dat gebruikers-adaptieve ondersteuning ook mogelijk is op basis van informatie afkomstig van voornamelijk neuronale activiteit,.

De gebruikte meettechniek, waarbij onderscheid gemaakt wordt tussen verschillende inspanningsniveaus door middel van de beschreven classificatietechniek, verschilt behoorlijk van de meer traditionele statistische aanpak, waardoor een goede vergelijking met andere meettechnieken lastig is. Zoals gezegd, de CSP procedure is een individuele aanpak waarmee verschillen tussen twee dataklassen (bijvoorbeeld twee inspanningsniveaus) gemaximaliseerd worden. Toch suggereren de zeer hoge classificatieprestaties dat deze techniek gevoeliger is voor veranderingen in mentale inspanning dan zowel de SDLP en de subjectieve maten.

### Hoofdstuk 7. Een op hersengolven en rijprestatie gebaseerde adaptieve cruise control

In dit onderzoek werden twee versies van een passief BCI systeem in een rijsimulator geïmplementeerd, gericht op het in stand houden van een comfortabel inspanningsniveau door de rijsnelheid aan te passen. De term passieve BCI wordt gebruikt om systemen aan te duiden waarbij hersenactiviteit wordt gekoppeld aan computeracties, echter op een onbewuste, impliciete manier. Dit in tegenstelling tot actieve BCI systemen waarbij gebruikers door middel van bewuste, vrijwillige gedachtecommando's een systeem proberen aan te sturen.

De belangrijkste reden om twee versies van een BCI te onderzoeken was om zowel een systeem te maken dat alleen reageert op neuronale activiteiten, als een systeem dat simpelweg de beste prestatie levert in termen van classificatie accuratesse, waarbij ook de informatie van mogelijk met spieractiviteit besmette kanalen en frequenties wordt gebruikt. Vanwege de bevinding dat frequenties in het EEG boven de 20 Hz vrijwel altijd mede bepaald worden door spieractiviteiten (Whitham et al., 2007), werden zowel een laagfrequentie als een hoogfrequentie BCI systeem getest. Er werd weer gereden op een rurale weg, waarbij bochtlengte werd gemanipuleerd om verschillende niveaus van stuur-moeilijkheid te creëren.

De adaptieve cruise-control, gebaseerd op EEG activiteit, moest drie niveaus van mentale inspanning kunnen onderscheiden om mentale inspanning constant te kunnen houden. Het systeem moest weten wanneer het tijd was om de rijsnelheid te verhogen in geval van weinig mentale inspanning, de rijsnelheid te verlagen in geval van veel mentale inspanning, en tot slot diende het te weten wanneer de rijsnelheid niet moest worden veranderd. Voor dit doel werd de binaire classificatie procedure die in hoofdstuk 6 beschreven staat uitgebreid naar een multipele classificatie procedure. Dit werd gedaan door middel van een paarsgewijze stemprocedure (Friedman, 1996). Dit houdt in dat voor aanvang van de daadwerkelijke applicatie van het BCI systeem, EEG data werd opgenomen uit drie korte ritten (dit wordt de kalibratiefase van het experiment genoemd). Deze ritten verschilden in de moeilijkheidsgraad en waren wederom geënt op het comfortabele inspanningsniveau, zoals bepaald in een initiële rit. De EEG data uit deze drie ritten vormden drie dataklassen (laag, comfortabel, en een hoog mentaal inspanningsniveau). Vervolgens werden per BCI versie, drie classificatiemodellen getraind; één model per dataklasse paar (laag vs. comfortabel, laag vs. hoog, comfortabel vs. hoog). Tijdens de daadwerkelijke toepassing (de applicatiefase van het experiment), werden alle drie de modellen gebruikt om waarschijnlijkheden toe te kennen aan binnenkomende, nieuwe EEG datasegmenten van één seconde. Deze waarschijnlijkheden gaven aan hoe zeker het model was of een datasegment het beste paste bij dataklasse A of B (bijvoorbeeld, 20% kans dat de deelnemer zwaar belast was tegenover 80% kans dat de deelnemer juist heel licht belast was). Met andere woorden, aan elk van de drie dataklassen werden twee waarschijnlijkheden toegekend, welke vervolgens opgeteld werden. De dataklasse die de hoogste gesommeerde waarschijnlijkheid ontving 'won' de stemming. Deze 'conclusie' van het BCI systeem resulteerde vervolgens in een snelheidsadvies aan de cruise control.

Uit voorzorg werden ook voertuigparameters gemonitord, waarbij gebruik werd gemaakt van een 'lopende' tijdperiode van twintig seconden. Hierdoor kon het protocol dat de rijsnelheid bepaalde, uitgebreid worden met interventies op basis van het rijgedrag. Ook werd in de simulator een absolute minimale en maximale rijsnelheid ingesteld. Als laatste werd na elke snelheidsverandering het systeem tien seconden 'op slot' gezet, waardoor snelle fluctuaties in de rijsnelheid voorkomen werden en de bestuurders de tijd werd gegund te wennen aan de nieuwe rijsnelheid.

De EEG data uit de kalibratiefase werden niet alleen gebruikt om de verschillende classificatiemodellen te trainen, maar ook om achteraf vast te stellen in hoeverre de resultaten uit het vorige onderzoek (hoofdstuk 6) gerepliceerd konden worden. Hieruit bleek dat de classificatieprestaties inderdaad vergelijkbaar waren. Voordat verder wordt gegaan met de bespreking van de resultaten uit de applicatiefase van deze studie, moet echter opgemerkt worden dat de dataclassificaties die door het adaptieve systeem gebruikt werden, waarschijnlijk enige vertraging hadden. Dat wil zeggen, tussen de tijd dat een EEG segment was gesampled en de tijd dat de uiteindelijke classificatie invloed had op de snelheidsbeslissing, zat voor een substantieel deel van de deelnemers één tot vijf seconden vertraging. Hoewel de impact van deze vertraging waarschijnlijk beperkt was doordat het systeem na elke verandering op slot ging, kan niet worden uitgegaan van een onmiddellijke koppeling tussen het momentane EEG en de rijsnelheid. Dit beperkt de conclusies die uit de applicatiefase van dit onderzoek getrokken kunnen worden.

Dit gezegd hebbende, bleek dat er tijdens de applicatiefase grote variaties waren in de manier hoe het systeem de rijsnelheid reguleerde, zowel tussen de deelnemers als ook tussen de verschillende condities binnen elke deelnemer. Deze grote verscheidenheid in (soms ook wispelturig) systeemgedrag, is een sterke indicatie dat de classificatieprestatie in de applicatiefase sterk teruggelopen was in vergelijking met de kalibratiefase. Het kwam bijvoorbeeld verschillende keren voor dat het BCI systeem telkens een snelheidsvermindering adviseerde, waardoor de rijsnelheid vrijwel de gehele rit op het absolute minimum lag. Het kwam ook regelmatig voor dat het BCI systeem altijd een snelheidstoename adviseerde. Hierdoor werd op deze basis de rijsnelheid vaak verhoogd, maar tegelijkertijd vaak weer verlaagd doordat voertuigparameters aangaven dat bestuurder de controle aan het verliezen was. Het beslisalgoritme gaf daarbij het advies op basis van rijgedrag altijd prioriteit boven het advies op basis van EEG activiteit.

Op basis van deze resultaten kan niet anders dan geconcludeerd worden dat er een grote afname was in classificatieprestatie tussen de kalibratie- en applicatiefase van de meeste experimentele sessies, wat het vermogen van het systeem om betrouwbaar mentale inspanning te monitoren ernstig heeft benadeeld. In de literatuur zijn tijd- en contextafhankelijkheid van fysiologische maten een bekend probleem, zowel binnen de BCI onderzoeksveld als binnen het bredere onderzoeksveld van mens-machine interactie (Shenoy et al., 2006; Fairclough, 2009). In lijn hiermee toont dit onderzoek aan dat substantiële verbeteringen nodig zijn, bijvoorbeeld door het classificatiemodel te blijven updaten tijdens de applicatiefase, voordat BCI technieken kunnen worden ingezet om mentale inspanning te reguleren tijdens de taakuitvoering.

## Hoofdstuk 8. Algemene discussie.

In de introductie werd gesteld dat dit proefschrift is gericht op de vooruitgang van mens-machine interactie. Hoe kunnen we iemands mentale toestand betrouwbaar monitoren en hoe kan met deze informatie de gebruiker ondersteund worden? Het is duidelijk dat hiervoor vele informatiebronnen gebruikt kunnen worden, maar ook dat geen van deze bronnen afzonderlijk een volledig, goed beeld geven. Met het onderzoek dat beschreven is in dit proefschrift zijn een aantal van deze opties bestudeerd. Door verschillende soorten rijgedrag te meten, door fysiologisch metingen te doen zoals als hartslag, ademhaling en hersenactiviteit, en als laatste door te vragen naar subjectieve ervaringen van de deelnemers werd onderzocht in hoeverre deze maten geschikt zijn voor gebruik door een gebruikers-adaptief systeem. Hierbij werd zowel gebruikt gemaakt van de traditionele statistische analyse methode, alsmede analyses uit te voeren op individueel niveau. De inzichten die hieruit voortvloeiden werden vervolgens gebruikt voor de creatie van twee adaptieve systemen. Kortom, dit proefschrift kan gezien worden als een stap in de ontwikkeling van gebruikers-adaptieve systemen. De resultaten uit de besproken onderzoeken laten zien dat, ten minste voor een systeem gebaseerd op fysiologische maten, het vervolgonderzoek zich zou moeten richten op het verbeteren van de monitorbetrouwbaarheid door het verlagen van de tijds- en contextafhankelijkheid. Bijvoorbeeld door verdere exploratie van de eerder genoemde data-gedreven data-analyse technieken. Tegelijkertijd moeten we blijven onderzoeken welke databronnen het meest informatief zijn om een preselectie van deze bronnen mogelijk te maken. Het is namelijk niet onwaarschijnlijk dat een puur, data-gedreven, mentale toestandsmonitor, die simpelweg zoveel mogelijk databronnen gebruikt, uiteindelijk te specifiek blijkt te zijn om van nut te zijn voor een breed scala van toepassingen.

# Dankwoord

# Acknowledgements

Het schrijven van een proefschrift en natuurlijk de jaren van onderzoek die daaraan vooraf gingen, zouden niet mogelijk zijn geweest zonder de ondersteuning van vele mensen in mijn omgeving, namelijk begeleiders, collega's , vrienden en familie. Hiervoor wil ik graag mijn dank uiten.

In de eerste plaats, of liever gezegd, in de eerste, tweede, en derde plaats wil ik Karel, Dick en Ben bedanken (in willekeurig volgorde!), die mij hebben aangenomen op het Europese REFLECT project. Ik kon altijd op jullie deur kloppen om zowel de grote lijnen als de details van het onderzoek te bespreken, elk op jullie eigen, en in mijn ogen aanvullende, manier. Ben, jouw grote en grondige interesse in onderzoek, o.a. met fysiologie, was altijd een inspiratie, en nadat je met pensioen bent gegaan heb ik ons veelvuldig contact zeer gemist. Karel, jij hebt me wetenschap als een 'way of life' laten zien. Ik herinner me bijvoorbeeld een opmerking tijdens een congresdiner in Zweden: 'Je moet er wel van genieten, niet vergeten hoor! ', wat in mijn beleving alles in het juiste perspectief zette en nog steeds zet. Dick, uiteindelijk heb ik het meest geleerd van jouw pragmatische insteek en oog voor detail. Zonder jouw steun waren de onderzoeken in dit proefschrift niet mogelijk geweest. Dank daarvoor. Als vierde wil ik op deze plek Ritske noemen. Hoewel het hebben van drie begeleiders al luxe was, vervulde ook jij deze rol tijdens de laatste fase van het onderzoek in dit proefschrift. Je enthousiasme en kennis heb ik steeds enorm gewaardeerd.

To properly thank my fellow PhD students who have made working life extra interesting, it is probably better to switch to English. To start with, a special thanks for my 'paranimfs' Ben and Arjan. You observed me being sleep deprived more than most people at work during those busy years, but you were always there to cheer me up with a joke or two, and it is very much appreciated in addition to all the work related discussions and good advises! Our ways have split up since, but I'm looking forward to meeting up every now and again for a very long time.

Since I started working at the RuG, I have shared an office with a number of people as our traffic group switched departments twice. To my roomies, and in order of appearance, Paolo, Berfu, Ellen, Janet, Eva, and Dafne, all of you thanks for making working in the office enjoyable and gratifying. Also Ari, Ebru, Frank (not yet(?) a PhD student, I know), Dina, Leon, Linda, Martijn, Roos, Stella, and Tita thanks for all the conversations and nice moments. You have made my PhD time in my city of birth a very international experience. Lastly, I'd like to thank the PhD students I worked with in the BCN PhD council and the 'promoraad' for the joint efforts to represent the PhD community, especially Nicole and Renate who were my predecessor and successor as chair of the promoraad.

Zonder de hulp van de collega's van de technische ondersteuning hadden de beoogde onderzoeken nooit gerealiseerd kunnen worden. Ik heb veel respect voor het werk dat jullie kunnen leveren. Met name Joop, Remco, Peter en Mark, dank voor jullie vele aanwijzingen en hulp bij de experimentele opstellingen en tijdens het leren programmeren.

I would also like to thank the partners in the REFLECT project, especially Marjolein, Christian, Gian Mario, Andrea, and Stephen Fairclough, thanks for the international research collaborations and good times we had during all the meetings. Furthermore, the reading committee, Marieke Martens, Monique Lorist and again Stephen Fairclough, thank you for agreeing to review and, of course, for approving my thesis.

Vrienden, het contact in de afgelopen jaren is niet zo veelvuldig geweest als gewenst. Maar weet dat ik deze gelegenheden altijd zeer gewaardeerd heb en er veel morele steun uit geput heb, juist in dezelfde jaren. Eduard, Klaas-Jan, Maurits, Niels, Peter, Sander, Willem, en ook Marjan, fijn dat jullie er waren. Dat geldt trouwens ook voor de buurt- en levensfasegenoten 'de Heertjes' en de familie Elzes-ten Hoopen. Ik hoop dat we nog lange tijd op zo'n prettige manier gezinsperikelen kunnen delen.

Ellen (Koog), ik wil jou als eerste familielid bedanken. Niet alleen omdat je in een drukke privéperiode toch de tijd hebt genomen om een deel van die proefschrift door te lezen en er foutjes uit te plukken, maar ook omdat je altijd iemand bent geweest om mee te praten. Verder wil ik voor alle goede zorgen, geduld, hulp en gezelligheid bedanken: broers, zwagers, schoonzussen en –ouders: Klaas-Pieter, Charissa, Heinse, Susanne, André, Remko, Roelof en Suze. En natuurlijk mijn ouders, Jitse en Riemy, jullie stonden altijd klaar om te helpen als dat nodig was met onze drukke privéleven, ook om mij de gelegenheid te geven aan dit proefschrift te werken. Pa en Ma, heel erg bedankt.

Tenslotte mijn lieve vrouw en kinderen. Jullie zijn het centrum van mijn bestaan en degenen voor wie ik het uiteindelijk allemaal doe. Ineke, je bent een absolute topper. Naast de drie kinderen die we in relatief korte tijd kregen met de daaraan gekoppelde drukte en je eigen werk ,vond je ook nog eens de ruimte en tijd om me te ondersteunen tijdens mijn promoveren, zelfs als ik vrije tijd opgaf in de avonden, weekenden, en soms zelfs vakantie om dit af te kunnen maken. Desondanks wilde je met me trouwen en vanaf vorig jaar kan ik met trots zeggen dat je mijn vrouw bent. Bedankt voor alles. Thieme, Hugo en Sterre, op het moment van schrijven zijn jullie 6, 4 en 2 jaar oud. Het is een cliché, maar absoluut waar, de tijd is ontzettend snel gegaan de afgelopen jaren en als ik naar jullie kijk besef ik dat dit promotietraject, ondanks alle moeite die er in ging zitten, slechts van ondergeschikt belang is. Jullie betekenen alle vier vele malen meer voor mij.

# References

Åkerstedt, T. (2004). Ambulatory EEG methods and sleepiness. In N. Stanton, A. Hedge, H. W. Hendrick, K. A. Brookhuis & E. Salas (Eds.), *Handbook of Ergonomics and Human Factors Methods,* London: Taylor & Francis.

Anund, A., Kecklund, G., Vadeby, A., Hjälmdahl, M., & Åkerstedt, T. (2008). The alerting effect of hitting a rumble strip—A simulator study with sleepy drivers. *Accident Analysis & Prevention, 40*(6), 1970-1976.

Audacity, (Version 1.2.4). [Computer program]. Available from: http://audacity.sourceforge.net/ [Accessed 6 February 2010].

Bainbridge, L. (1983). Ironies of automation. *Automatica, 19*(6), 775-779.

Banks, S. B., & Lizza, C. S. (1991). Pilot's Associate: A cooperative, knowledge-based system application. *IEEE expert, 6*(3), 18-29.

Bella, F. (2008). Driving simulator for speed research on two-lane rural roads. *Accident Analysis & Prevention, 40*(3), 1078-1087.

Berlyne, D. E. (1971). *Aesthetics and psychobiology.* New York: Appleton-Century, Crofts.

Billings, C. E. (1997). *Aviation automation: The search for a human-centered approach.* Mahwah, NJ: Erlbaum.

Blaauw, G. J. (1982). Driving experience and task demands in simulator and instrumented car: a validation study. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 24*(4), 473-486.

Blacking, J. (1973). *How musical is man?* Seattle, WA: University of Washington Press.

Blankertz, B., Tomioka, R., Lemm, S., Kawanabe, M., & Muller, K. R. (2008). Optimizing spatial filters for robust EEG single-trial analysis. *Signal Processing Magazine, IEEE, 25*(1), 41-56.

Boiten, F. A., Frijda, N. H., & Wientjes, C. J. (1994). Emotions and respiratory patterns: review and critical analysis. *International Journal of Psychophysiology, 17*(2), 103-128.

Borghini, G., Astolfi, L., Vecchiato, G., Mattia, D., & Babiloni, F. (In press. 2012). Measuring neurophysiological signals in aircraft pilots and car drivers for the assessment of mental workload, fatigue and drowsiness. Neuroscience & Biobehavioral Reviews.

Boucsein, W., & Backs, R. W. (2000). Engineering psychophysiology as a discipline: Historical and theoretical aspects. In: R.W.Backs, W. Boucsein, R. W. Backs & W. Boucsein (Eds.), Engineering psychophysiology: Issues and applications (pp. 3-30), Mahwah, NJ US: Lawrence

Erlbaum Associates Publishers.

Brehmer, B. (1990). Variable errors set a limit to adaptation. Ergonomics, 33(10-11), 1231-1239.

Brodie, L., Lyndal, B., & Elias, I. J. (2009). Heavy vehicle driver fatalities: Learning's from fatal road crash investigations in Victoria. Accident Analysis & Prevention, 41(3), 557-564.

Brodsky, W. (2001). The effects of music tempo on simulated driving performance and vehicular control. Transportation research part F: traffic psychology and behaviour, 4(4), 219-241.

Brookhuis, K. A., & de Waard, D. (2007). Intelligent transport systems for vehicle drivers. In T. Gärling & E. M. Steg (Eds.), Threats from car traffic (pp. 383-401), Amsterdam, Netherlands: Elsevier.

Brookhuis, K. A., & de Waard, D. (2010). Monitoring drivers' mental workload in driving simulators using physiological measures. Accident Analysis & Prevention, 42(3), 898-903.

Brookhuis, K. A., de Waard, D., & Fairclough, S. H. (2003). Criteria for driver impairment. Ergonomics, 46(5), 433-445.

Brookhuis, K. A., de Waard, D., & Samyn, N. (2004). Effects of MDMA (ecstasy), and multiple drugs use on (simulated) driving performance and traffic safety. Psychopharmacology, 173(3-4), 440-445.

Broström, R., Engström, J., Agnvall, A., & Markkula, G. (2006). Towards the next generation intelligent driver information system (IDIS): The Volvo car interaction manager concept. Proceedings of the 13th ITS world congress, London, 8-12 October, 2006.

Brouwer, A. M., Hogervorst, M. A., Van Erp, J. B., Heffelaar, T., Zimmerman, P. H., & Oostenveld, R. (2012). Estimating workload using EEG spectral power and ERPs in the n-back task. Journal of Neural Engineering, 9(4), 045008.

Brusilovsky, P. (2001). Adaptive hypermedia. User modeling and user-adapted interaction, 11(1-2), 87-110.

Brusilovsky, P., & Millán, E. (2007). User models for adaptive hypermedia and adaptive educational systems. In P. Brusilovsky, A. Kobsa & W. Nejdl (Eds.), The adaptive web (pp. 3-53), Berlin, Heidelberg, New York: Springer Verlag.

Byrne, E. A., & Parasuraman, R. (1996). Psychophysiology and adaptive automation. Biological psychology, 42(3), 249-268.

Cacioppo, J. T., & Gardner, W. L. (1999). Emotion. Annual review of psychology, 50(1), 191-214.

Cacioppo, J. T., Tassinary, L. G., Berntson, G. G. (2000). Psychophysiological science. In J. T. Cacioppo, L. G. Tassinary & G. G. Berntson (Eds.), Handbook of Psychophysiology, (pp. 3-26), Cambridge, UK: Cambridge University Press.

Carsten, O. M. J., & Brookhuis, K. (2005). Issues arising from the HASTE experiments. Transportation Research Part F: Traffic Psychology and Behaviour, 8(2), 191-196.

Christensen, J. C., & Estepp, J. R. (2013). Coadaptive Aiding and Automation Enhance Operator Performance. Human Factors: The Journal of the Human Factors and Ergonomics Society, 55(5), 965-975.

Chu, Y. Y., & Rouse, W. B. (1979). Adaptive allocation of decisionmaking responsibility between human and computer in multitask situations. Systems, Man and Cybernetics, IEEE Transactions on, 9(12), 769-778.

Clarke, D. D., Ward, P., Bartle, C., & Truman, W. (2010). Killer crashes: Fatal road traffic accidents in the UK. Accident Analysis & Prevention, 42(2), 764-770.

Constantinou, E., Panayiotou, G., Konstantinou, N., Loutsiou-Ladd, A., & Kapardis, A. (2011). Risky and aggressive driving in young adults: Personality matters. Accident Analysis & Prevention, 43(4), 1323-1331.

Cutrell, E., & Tan, D. (2008). BCI for passive input in HCI. In Proceedings of CHI (Vol. 8, pp. 1-3).

Dalton, B. H., & Behm, D. G. (2007). Effects of noise and music on human and task performance: A systematic review. Occupational Ergonomics, 7(3), 143-152.

Davis, C. A. (1992). The effects of music and basic relaxation instruction on pain and anxiety of women undergoing in-office gynecological procedures. Journal of Music Therapy, 29, 202-202.

de Jong, M. A., Van Mourik, K. R., & Schellekens, H. M. C. (1973). A physiological approach to aesthetic preference. Psychotherapy and Psychosomatics, 22(1), 46-51.

de Waard, D. (1996). The measurement of drivers' mental workload (Doctoral dissertation). University of Groningen, Traffic Research Center, Haren, Netherlands.

de Waard, D., & Brookhuis, K. A. (1991). Assessing driver status: a demonstration experiment on the road. Accident analysis & prevention, 23(4), 297-307.

de Waard, D., Brookhuis, K. A., & Hernandez-Gress, N. (2001). The feasibility of detecting phone-use related driver distraction. International Journal of Vehicle Design, 26(1), 85-95.

de Waard, D., Dijksterhuis, C., & Brookhuis, K. A. (2009). Merging into heavy motorway traffic by young and elderly drivers. Accident Analysis & Prevention,41(3), 588-597.

de Waard, D., Jessurun, M., Steyvers, F. J. J. M., Raggatt, P. T. F., & Brookhuis, K. A. (1995). Effect of road layout and road environment on driving performance, drivers' physiology and road appreciation. Ergonomics, 38(7), 1395-1407.

de Waard, D., Kruizinga, A., & Brookhuis, K. A. (2008). The consequences of an increase in heavy goods vehicles for passenger car drivers' mental workload and behaviour: a simulator study. Accident Analysis & Prevention, 40(2), 818-828.

de Waard, D., Steyvers, F. J., & Brookhuis, K. A. (2004). How much visual road information is needed to drive safely and comfortably? Safety Science,42(7), 639-655.

de Waard, D., van der Hulst, M., Hoedemaeker, M., & Brookhuis, K. A. (1999). Driver behavior in an emergency situation in the Automated Highway System. Transportation human factors, 1(1), 67-82.

DeNora, T. (2000). Music in everyday life. Cambridge: Cambridge University Press.

DeNora, T. (2003). Music sociology: Getting the music into the action. British Journal of Music Education, 20(2), 165-177.

Dibben, N., & Williamson, V. J. (2007). An exploratory survey of in-vehicle music listening. Psychology of Music, 35(4), 571-589.

Dijksterhuis, C., Brookhuis, K. A., & de Waard, D. (2011). Effects of steering demand on lane keeping behaviour, self-reports, and physiology. A simulator study. Accident Analysis & Prevention, 43(3), 1074-1081.

Dijksterhuis, C., de Waard, D., Brookhuis, K. A., Mulder, B. L., & de Jong, R. (2013). Classifying visuomotor workload in a driving simulator using subject specific spatial brain patterns. Frontiers in neuroscience, 7.

Dijksterhuis, C., Stuiver, A., Mulder, B., Brookhuis, K. A., & de Waard, D. (2012). An Adaptive Driver Support System User Experiences and Driving Performance in a Simulator. Human Factors: The Journal of the Human Factors and Ergonomics Society, 54(5), 772-785.

Dornhege, G., Blankertz, B., Curio, G., & Muller, K. R. (2004). Boosting bit rates in noninvasive EEG single-trial classifications by feature combination and multiclass paradigms. Biomedical Engineering, IEEE Transactions on, 51(6), 993-1002.

Duric, Z., Gray, W. D., Heishman, R., Li, F., Rosenfeld, A., Schoelles, M. J., ... & Wechsler, H. (2002). Integrating perceptual and cognitive modeling for adaptive and intelligent human-computer interaction. Proceedings of the IEEE, 90(7), 1272-1289.

Endsley, M. R. (1995). Toward a theory of situation awareness in dynamic systems. Human Factors: The Journal of the Human Factors and Ergonomics Society, 37(1), 32-64.

Endsley, M. R. (1999). Level of automation effects on performance, situation awareness and workload in a dynamic control task. Ergonomics, 42(3), 462-492.

European Commission. (2013). eCall: automated emergency call for road accidents mandatory in cars from 2015 [Press release]. Retrieved from http://europa.eu/rapid/press-release_IP-13-534_en.htm

Fadden, S., Ververs, P. M., & Wickens, C. D. (2001). Pathway HUDs: are they viable?. Human Factors: The Journal of the Human Factors and Ergonomics Society, 43(2), 173-193.

Fairclough, S. H. (2009). Fundamentals of physiological computing. Interacting with computers, 21(1), 133-145.

Feigh, K. M., Dorneich, M. C., & Hayes, C. C. (2012). Toward a Characterization of Adaptive Systems A Framework for Researchers and System Designers. Human Factors: The Journal of the Human Factors and Ergonomics Society, 54(6), 1008-1024.

Flemisch, F., Kelsch, J., Löper, C., Schieben, A., & Schindler, J. (2008). Automation spectrum, inner/outer compatibility and other potentially useful human factor concepts for assistance. and automation. In D. de Waard, F. O. Flemisch, B. Lorenz, H. Oberheid & K. A. Brookhuis (Eds.), Human factors for assistance and automation (pp. 257-272), Maastricht, Netherlands: Shaker.

Friedman, J. (1996). Another approach to polychotomous classifcation. Technical report, Stanford University, Department of Statistics.

Fries, P., Nikolić, D., & Singer, W. (2007). The gamma cycle. Trends in neurosciences, 30(7), 309-316.

Fuller, R. (2005). Towards a general theory of driver behaviour. Accident Analysis & Prevention, 37(3), 461-472.

Gaillard, A. W. K., & Kramer, A. F. (2000). Theoretical and methodological issues in psychophysiological research. In R. W. Backs & W. Boucsein (Eds.), Engineering psychophysiology: Issues and applications (pp. 31-58), Mahwah, NJ: Lawrence Erlbaum.

Gish, K. W., & Staplin, L. (1995). Human factors aspects of using head-up displays in automobiles: A review of the literature (Report DOT HS 808 320). Washington, DC: U.S. Department of Transportation.

Godley, S. T., Triggs, T. J., & Fildes, B. N. (2002). Driving simulator validation for speed research. Accident analysis & prevention, 34(5), 589-600.

Godley, S. T., Triggs, T. J., & Fildes, B. N. (2004). Perceptual land width, wide perceptual road centre markings and driving speeds. Ergonomics, 47(3), 237-257.

Gomez, P., & Danuser, B. (2004). Affective and physiological responses to environmental noises and music. International Journal of Psychophysiology,53(2), 91-103.

Gratton, G., Coles, M. G., & Donchin, E. (1983). A new method for off-line removal of ocular artifact. Electroencephalography and clinical neurophysiology,55(4), 468-484.

Grosse-Wentrup, M., & Buss, M. (2008). Multiclass common spatial patterns and information theoretic feature extraction. Biomedical Engineering, IEEE Transactions on, 55(8), 1991-2000.

Grossman, P., & Taylor, E. W. (2007). Toward understanding respiratory sinus arrhythmia: relations to cardiac vagal tone, evolution and biobehavioral functions. Biological psychology, 74(2), 263-285.

Hancock, P. A. (1989). A dynamic model of stress and sustained attention. Human Factors: The Journal of the Human Factors and Ergonomics Society,31(5), 519-537.

Hancock, P. A., & Chignell, M. H. (1988). Mental workload dynamics in adaptive interface design. Systems, Man and Cybernetics, IEEE Transactions on, 18(4), 647-658.

Hancock, P. A., & Verwey, W. B. (1997). Fatigue, workload and adaptive driver systems. Accident Analysis & Prevention, 29(4), 495-506.

Hargreaves, D., & North, A. (2010). Experimental aesthetics and liking for music. In P. Juslin & J. Sloboda (Eds.), Handbook of music and emotions: theory, research, applications (pp. 515-547), Oxford: Oxford University Press.

Hettinger, L. J., Branco, P., Encarnacao, L. M., & Bonato, P. (2003). Neuroadaptive technologies: applying neuroergonomics to the design of advanced interfaces. Theoretical Issues in Ergonomics Science, 4(1-2), 220-237.

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(1), 73-93.

Hockey, G. R. J. (2003). Operator functional state as a framework for the assessment of performance degradation. In G. R. J. Hockey, A. W. K. Gaillard & O. Burov (Eds.), Operator functional state (pp. 8-23), Amsterdam: IOS Press.

Hockey, G. R. J., Nickel, P., Roberts, A. C., & Roberts, M. H. (2009). Sensitivity of candidate markers of psychophysiological strain to cyclical changes in manual control load during simulated process control. Applied Ergonomics, 40(6), 1011-1018.

Homma, I., & Masaoka, Y. (2008). Breathing rhythms and emotions. Experimental physiology, 93(9), 1011-1021.

Hoogeboom, P. J., & Mulder, L. J. M. (2004). Physiological indices for the estimation of momentary changes in cognitive work-load and mental state. In D. de Waard, K. A. Brookhuis & C. M. Weikert (Eds.), Human factors in design (pp. 147-160), Maastricht, Netherlands: Shaker.

Horrey, W. J., & Wickens, C. D. (2004). Driving and side task performance: The effects of display clutter, separation, and modality. Human Factors: The Journal of the Human Factors and Ergonomics Society, 46(4), 611-624.

Inagaki, T. (2003). Adaptive automation: Sharing and trading of control. In E. Hollnagel (Ed.), Handbook of cognitive task design (pp. 221-245), Mahwah, NJ: Erlbaum.

Iwanaga, M., & Moroki, Y. (1999). Subjective and physiological responses to music stimuli controlled over activity and preference. Journal of Music Therapy,36, 26-38.

Jamson, S., Wardman, M., Batley, R., & Carsten, O. (2008). Developing a driving Safety Index using a Delphi stated preference experiment. Accident Analysis & Prevention, 40(2), 435-442.

Johnson, A. K., & Anderson, E. A. (1990). Stress and arousal. In J. T. Cacioppo, & L. G. Tassinary, Principles of psychophysiology (pp. 216-252). Cambridge: Cambridge University Press.

Juslin, P. N., & Sloboda, J. A. (Eds.) (2010). Handbook of music and emotion: Theory, research, applications, Oxford, Oxford University Press.

Kaber, D. B., & Endsley, M. R. (2004). The effects of level of automation and adaptive automation on human performance, situation awareness and workload in a dynamic control task. Theoretical Issues in Ergonomics Science, 5(2), 113-153.

Kaber, D. B., & Prinzel III, L. J. (2006). Adaptive and Adaptable Automation Design: A Critical Review of the Literature and Recommendations for Future Research. Report No. NASA/TM-2006-214504. Retrieved from http://ntrs.nasa.gov/search.jsp?R=20060053373

Kaber, D. B., & Riley, J. M. (1999). Adaptive automation of a dynamic control task based on secondary task workload measurement. International journal of cognitive ergonomics, 3(3), 169-187.

Kahneman, D. (1973). Attention and effort. New Jersey, U.S.A.: Prentice Hall.

Knight, W. E., & Rickard, N. S. (2001). Relaxing music prevents stress-induced increases in subjective anxiety, systolic blood pressure, and heart rate in healthy males and females. Journal of music therapy, 38(4), 254-272.

Kohlmorgen J., Dornhege G., Braun M., Blankertz B., Müller K. R., Curio G., Hagemann K., Bruns A., Schrauf M., & Kincses W. (2007). Improving human performance in a real operating environment through real-time mental workload detection. In G. Dornhege, J. d. R. Millán, T. Hinterberger, D. McFarland & K. R. Müller (Eds.), Toward Brain-Computer Interfacing (pp. 409-422), Cambridge, MA: MIT press.

Konecni, V. J. (1982). Social interaction and musical preference. In D. Deutsch (Ed.), The psychology of music (pp. 297-516), New York: Academic Press.

Krumhansl, C. L. (1997). An exploratory study of musical emotions and psychophysiology. Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale, 51(4), 336.

Lee, J. D. (2008). Review of a pivotal Human Factors article: "Humans and automation: use, misuse, disuse, abuse". Human Factors: The Journal of the Human Factors and Ergonomics Society, 50(3), 404-410.

Lee, J. D. (2010, October). Making the human-technology marriage work. Keynote presented at the conference of the Human Factors and Ergonomics Europe chapter, Berlin, Germany.

Lewis-Evans, B. (2012). Testing models of driver behaviour (Doctoral dissertation), Groningen: University Library [Host].

Lewis-Evans, B., & Charlton, S. G. (2006). Explicit and implicit processes in behavioural adaptation to road width. Accident Analysis & Prevention, 38(3), 610-617.

Lewis-Evans, B., de Waard, D., & Brookhuis, K. A. (2011). Speed maintenance under cognitive load-Implications for theories of driver behaviour. Accident Analysis & Prevention, 43(4), 1497-1507.

Lewis-Evans, B., & Rothengatter, T. (2009). Task difficulty, risk, effort and comfort in a simulated driving task—Implications for Risk Allostasis Theory.Accident Analysis & Prevention, 41(5), 1053-1063.

Markoff, J. (2010, October 9). Google's cars drive themselves, in traffic. The New York Times. Retrieved from http://www.nytimes.com/2010/10/10/science/10google.html

Matthews, G., & Desmond, P. A. (2002). Task-induced fatigue states and simulated driving performance. The Quarterly Journal of Experimental Psychology: Section A, 55(2), 659-686.

Matthews, G., Jones, D. M., & Chamberlain, A. G. (1990). Refining the measurement of mood: the UWIST Mood Adjective Checklist. British journal of psychology, 81(1), 17-42.

Michon, J. A. (1985). A critical view of driver behavior models: What do we know, what should we do? In L. Evans & R. C. Schwing (Eds.), Human behavior and traffic safety (pp. 485-524), New York, NY: Plenum Press.

Michon, J. A. (Ed.) (1993). Generic intelligent driver support, London: Taylor and Francis.

Miller, C.A., & Funk, H.B. (2001). Associates with etiquette: Meta-communications to make human-automation interaction more natural, productive and polite. In proceedings of the 8th European Conference on Cognitive Science Approaches to Process Control, Munich, September 24-26, 2001.

Miller, C. A., & Hannen, M. D. (1999). The Rotorcraft Pilot's Associate: design and evaluation of an intelligent user interface for cockpit information management. Knowledge-Based Systems, 12(8), 443-456.

Miller, C. A., & Parasuraman, R. (2007). Designing for flexible interaction between humans and automation: Delegation interfaces for supervisory control. Human Factors: The Journal of the Human Factors and Ergonomics Society,49(1), 57-75.

Morrison, J. G., & Gluckman, J. P. (1994). Definitions and prospective guidelines for the application of adaptive automation. In M. Mouloua & R. Parasuraman (Eds.), Human performance in automated systems: Current research and trends (pp. 256-263), Hillsdale, NJ: Lawrence Erlbaum.

Mulder, G. (1986). The concept and measurement of mental effort. In G. R. J. Hockey, A. W. K. Gaillard & M.G.H. Coles (Eds.), Energetics and human information processing (pp. 175-198), Dordrecht, The Netherlands: Martinus Nijhoff Publishers.

Mulder, L. J. M. (1992). Measurement and analysis methods of heart rate and respiration for use in applied environments. Biological psychology, 34(2), 205-236.

Mulder, L. J. M., de Waard, D., Hoogenboom, P. J., Quispel, L., & Stuiver, A. (2008). Using physiological measures for task performance: Towards a companion. In J. H. D. M. Westerink, M. Ouwerkerk, T. J. M. Overbeek, W. F. Pasveer & B. de Ruyter (Eds.), Probing experience (pp. 221-234), Dordrecht, Netherlands: Springer.

Mulder, L. J. M., Dijksterhuis, C., Stuiver, A., & de Waard, D. (2009). Cardiovascular state changes during performance of a simulated ambulance dispatchers' task: Potential use for adaptive support. Applied ergonomics, 40(6), 965-977.

Müller, K. R., Tangermann, M., Dornhege, G., Krauledat, M., Curio, G., & Blankertz, B. (2008). Machine learning for real-time single-trial EEG-analysis: from brain-computer interfacing to mental state monitoring. Journal of neuroscience methods, 167(1), 82-90.

Muthukumaraswamy, S. D., & Singh, K. D. (2013). Visual gamma oscillations: The effects of stimulus type, visual field coverage and stimulus motion on MEG and EEG recordings. NeuroImage, 69, 223-230.

Näätänen, R., & Summala, H. (1976). Road-user behaviour and traffic accidents. Publication of: North-Holland Publishing Company.

North, A. C., & Hargreaves, D. J. (1999). Music and driving game performance.Scandinavian Journal of Psychology, 40(4), 285-292.

North, A. C., & Hargreaves, D. J. (2008). The social and applied psychology of music (pp. 81-239), Oxford: Oxford University Press.

Nyklíček, I., Thayer, J. F., & Van Doornen, L. J. (1997). Cardiorespiratory differentiation of musically-induced emotions. Journal of Psychophysiology, 11, 304-321.

O'Hanlon, J. F., Haak, T. W., Blaauw, G. J., & Riemersma, J. B. (1982). Diazepam impairs lateral position control in highway driving. Science,217(4554), 79-81.

Parasuraman, R., Bahri, R., Deaton, T., Morrison, J., & Barnes, M. (1992). Theory and design of adaptive automation in aviation systems (Progress Report No. NAWCADWAR-92033-60). Warminster, PA: Naval Air Warfare Center, Aircraft Division.

Parasuraman, R., & Manzey, D. H. (2010). Complacency and bias in human use of automation: An attentional integration. Human Factors: The Journal of the Human Factors and Ergonomics Society, 52(3), 381-410.

Parasuraman, R., & Riley, V. (1997). Humans and automation: Use, misuse, disuse, abuse. Human Factors: The Journal of the Human Factors and Ergonomics Society, 39(2), 230-253.

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

Parkes, A. M. (1991). Data capture techniques for RTI usability evaluation. In Proceedings of the DRIVE Conference (pp. 1440-1456). Amsterdam, Netherlands: Elsevier.

Pechacek, J., & Webb, R. (1996). Crewman's associate (CA) cognitive decision aiding (CDA) applications for the US Army's future main battle tanks. Expert Systems with Applications, 11(2), 191-206.

Pêcher, C., Lemercier, C., & Cellier, J. M. (2009). Emotions drive attention: Effects on driver's behaviour. Safety Science, 47(9), 1254-1259.

Peng, Y., Boyle, L. N., & Hallmark, S. L. (2012). Driver's lane keeping ability with eyes off road: Insights from a naturalistic study. Accident Analysis & Prevention.

Piferi, R. L., Kline, K. A., Younger, J., & Lawler, K. A. (2000). An alternative approach for achieving cardiovascular baseline: Viewing an aquatic video.International Journal of Psychophysiology, 37(2), 207-217.

Pope, A. T., Bogart, E. H., & Bartolome, D. S. (1995). Biocybernetic system evaluates indices of operator engagement in automated task. Biological psychology, 40(1), 187-195.

Prinzel III, L. J. (2002). Research on hazardous states of aware-ness and physiological factors in aerospace operations(Report No. NASA/TM-2002-211444). Hampton, VA: National Aeronautics and Space Administration.

Prinzel III, L. J., Freeman, F. G., Scerbo, M. W., Mikulka, P. J., & Pope, A. T. (2003). Effects of a psychophysiological system for adaptive automation on performance, workload, and the event-related potential P300 component.Human Factors: The Journal of the Human Factors and Ergonomics Society,45(4), 601-614.

Prinzel III, L. J., Pope, A. T., Freeman, F. G., Scerbo, M. W., & Mikulka, P. J. (2001). Empirical analysis of EEG and ERPs for psychophysiological adaptive task allocation. Report No. NASA/TM-2001-211016. Hampton, VA: NASA. Retrieved from http://nix.nasa.gov/search.jsp?R=20010060403

Ramaekers, J. G., Robbe, H. W. J., & O'Hanlon, J. F. (2000). Marijuana, alcohol and actual driving performance. Human Psychopharmacology: Clinical and Experimental, 15(7), 551-558.

Ranney, T. A. (1994). Models of driving behavior: a review of their evolution.Accident Analysis & Prevention, 26(6), 733-750.

Räsänen, M. (2005). Effects of a rumble strip barrier line on lane keeping in a curve. Accident Analysis & Prevention, 37(3), 575-581.

Recarte, M. A., & Nunes, L. M. (2000). Effects of verbal and spatial-imagery tasks on eye fixations while driving. Journal of Experimental Psychology: Applied, 6(1), 31.

Rickard, N. S. (2004). Intense emotional responses to music: A test of the physiological arousal hypothesis. Psychology of Music, 32(4), 371-388.

Ritz, T. (2004). Probing the psychophysiology of the airways: Physical activity, experienced emotion, and facially expressed emotion. Psychophysiology, 41(6), 809-821.

Rosey, F., Auberlet, J. M., Moisan, O., & Dupré, G. (2009). Impact of Narrower Lane Width. Transportation Research Record: Journal of the Transportation Research Board, 2138(1), 112-119.

Rouse, W. B. (1976). Adaptive allocation of decision making responsibility between supervisor and computer. In T.B. Sheridan & G. Johannsen (Eds.), Monitoring behavior and supervisory control (pp. 295-306), New York: Plenum Press.

Rouse, W. B. (1977). Human-computer interaction in multitask situations.Systems, Man and Cybernetics, IEEE Transactions on, 7(5), 384-392.

Rouse, W. B. (1988). Adaptive aiding for human/computer control. Human Factors: The Journal of the Human Factors and Ergonomics Society, 30(4), 431-443.

Rouse, W. B., Geddes, N. D., & Hammer, J. M. (1990). Computer-aided fighter pilots. Spectrum, IEEE, 27(3), 38-41.

Rouse, W. B., & Rouse, S. H. (1983). A framework for research on adaptive decision aids. ALPHATECH INC BURLINGTON MA.

Saarikallio, S., & Erkkilä, J. (2007). The role of music in adolescents' mood regulation. Psychology of Music, 35(1), 88-109.

Scerbo, M. W. (1996). Theoretical perspectives on adaptive automation. In R. Parasuraman & M. Mouloua (Eds.), Automation and human performance: Theory and applications (pp. 37-63), Hillsdale, NJ: Lawrence Erlbaum.

Scerbo, M. W. (2001). Adaptive automation. In W. Karwowski (Ed.), International encyclopedia of ergonomics and human factors (pp. 1077-1079), London, UK: Taylor and Francis.

Scerbo, M. W., Freeman, F. G., & Mikulka, P. J. (2003). A brain-based system for adaptive automation. Theoretical Issues in Ergonomics Science, 4(1-2), 200-219.

Schoffelen, J. M., Oostenveld, R., & Fries, P. (2005). Neuronal coherence as a mechanism of effective corticospinal interaction. Science, 308(5718), 111-113.

Serbedzija, N. B., & Fairclough, S. H. (2009, May). Biocybernetic loop: from awareness to evolution. In Evolutionary Computation, 2009. CEC'09. IEEE Congress on (pp. 2063-2069). IEEE.

Shek, V., & Schubert, E. (2009, December). Background Music at Work-A literature review and some hypotheses. In Proceedings of the 2nd international conference on music communication science (ICoMCS2) (pp. 87-91), Sydney, Australia, 3-4 December, 2009.

Shenoy, P., Krauledat, M., Blankertz, B., Rao, R. P., & Müller, K. R. (2006). Towards adaptive classification for BCI. Journal of neural engineering, 3(1), R13.

Sheridan, T. B. (1976a). Preview of models of the human monitor/supervisor. In T. B. Sheridan & G. Johannsen (Eds.), Monitoring behavior and supervisory control (pp. 175-180), New York: Plenum.

Sheridan, T. B. (1976b). Toward a general model of supervisory control. In T. B. Sheridan & G. Johannsen (Eds.), Monitoring behavior and supervisory control (pp. 271-281), New York: Plenum.

Sheridan, T. B. (2000). Function allocation: Algorithm, alchemy or apostasy? International Journal of Human Computer Studies, 52(2), 203-216.

Sheridan, T. B., & Parasuraman, R. (2005). Human-automation interaction.Reviews of human factors and ergonomics, 1(1), 89-129.

Sheridan, T. B., & Verplank, W. L. (1978). Human and computer control of undersea teleoperators (Technical Report). Cambridge, MA: MIT, Man-Machine Systems Laboratory, Department of Mechanical Engineering. Retrieved from http://www.dtic.mil/cgi-bin/GetTRDoc?AD=ADA057655

Skowronek, J., McKinney, M. F., & Van De Par, S. (2006, October). Ground truth for automatic music mood classification. In Proceedings of the 7th international conference on music information retrieval (pp. 395-396), Victoria, Canada: University of Victoria.

Skowronek, J., McKinney, M. F., & Van De Par, S. (2007). A demonstrator for automatic music mood estimation. In ISMIR (pp. 345-346). Vienna, Austria: Austrian Computer Society.

Sloboda, J. A., & Juslin, P. N. (2010). At the interface between the inner and outer world, psychological perspectives. In P. N. Juslin & J. A. Sloboda (Eds.), Handbook of music and emotion: theory, research, and applications (pp. 73-98), New York: Oxford University Press.

Staubach, M. (2009). Factors correlated with traffic accidents as a basis for evaluating Advanced Driver Assistance Systems. Accident Analysis & Prevention, 41(5), 1025-1033.

Stern, R. M., Ray, W. J., & Quigley, K. S. (2001). Psychophysiological recording. Oxford University Press.

Stevens, J. P. (2002). Applied Multivariate Statistics for the Social Sciences, Mahwaw, New Jersey: Lawrence Erlbaum Associates, Inc.

Strayer, D. L., Cooper, J. M., Turrill, J., Coleman, J., Medeiros-Ward, N., & Biondi, F. (2013). Measuring cognitive distraction in the automobile, Washington, DC: AAA Foundation for Traffic Safety Report.

Summala, H. (2005). Traffic psychology theories: Towards understanding driving behaviour and safety efforts. In G. Underwood (Ed.), Traffic and Transport Psychology (pp. 383-394), Amsterdam: Elsevier.

Summala, H. (2007). Towards understanding motivational and emotional factors in driver behaviour: Comfort through satisficing. In C. Cacciabue (Ed.), Modelling Driver Behaviour in Automotive Environments (pp. 189-207), London: Springer.

Summala, H., Nieminen, T., & Punto, M. (1996). Maintaining lane position with peripheral vision during in-vehicle tasks. Human Factors: The Journal of the Human Factors and Ergonomics Society, 38(3), 442-451.

Tattersall, A. J., & Fairclough, S. H. (2003). Adaptive automation and modes of control. In G. R. J. Hockey, A. W. K. Gaillard & O. Burov (Eds.), Operator functional state (pp. 238-248), Amsterdam: IOS Press.

Thomas, J. A., & Walton, D. (2007). Measuring perceived risk: Self-reported and actual hand positions of SUV and car drivers. Transportation research part F: traffic psychology and behaviour, 10(3), 201-207.

Thrun, S. (2010, October 10). What we're driving at[Blog post]. Retrieved from http://googleblog.blogspot.com/2010/10/what-were-driving-at.html

Ting, C. H., Mahfouf, M., Nassef, A., Linkens, D. A., Panoutsos, G., Nickel, P., ... & Hockey, G. (2010). Real-time adaptive automation system based on identification of operator functional state in simulated process control operations. Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on, 40(2), 251-262.

Törnros, J. (1998). Driving behaviour in a real and a simulated road tunnel—a validation study. Accident Analysis & Prevention, 30(4), 497-503.

Tufano, D. R. (1997). Automotive HUDs: The overlooked safety issues. Human Factors: The Journal of the Human Factors and Ergonomics Society, 39(2), 303-311.

UNECE, 2007. Statistics of Road Traffic Accidents in Europe and North America, 51th edition, Geneva, pp. 46-61.

United Nations, Economic Commission for Europe. (2007). Statistics of road traffic accidents in Europe and North America. Retrieved from http://www.unece.org/trans/main/wp6/trans statpub.html

van den Broek, E. L., & Westerink, J. H. (2009). Considerations for emotion-aware consumer products. Applied Ergonomics, 40(6), 1055-1064.

van der Laan, J. D., Heino, A., & de Waard, D. (1997). A simple procedure for the assessment of acceptance of advanced trans-port telematics. Transportation Research-Part C: Emerging Technologies, 5, 1-10.

van der Zwaag, M. D., Dijksterhuis, C., de Waard, D., Mulder, B. L., Westerink, J. H., & Brookhuis, K. A. (2012). The influence of music on mood and performance while driving. Ergonomics, 55(1), 12-22.

van der Zwaag, M. D., Westerink, J. H., & van den Broek, E. L. (2011). Emotional and psychophysiological responses to tempo, mode, and percussiveness. Musicae Scientiae, 15(2), 250-269.

van Driel, C. J., Davidse, R. J., & van Maarseveen, M. F. (2004). The effects of an edgeline on speed and lateral position: a meta-analysis. Accident Analysis & Prevention, 36(4), 671-682.

van Winsum, W., & van Wolffelaar, P.C. (1993). GIDS small world simulation. In: J.A. Michon (Ed.), Generic intelligent driver support (pp. 175-191), London: Taylor and Francis.

Walsh, M. J. (2010). Driving to the beat of one's own hum: automobility and musical listening. Studies in Symbolic Interaction, 35, 201-221.

Wang, Z., Hope, R. M., Wang, Z., Ji, Q., & Gray, W. D. (2012). Cross-subject workload classification with a hierarchical Bayes model. NeuroImage, 59(1), 64-69.

Weintraub, D. J. (1992). Human factors issues in head-up display design: The book of HUD (No. SOAR-92-2). DAYTON UNIV OH RESEARCH INST.

Whitham, E. M., Pope, K. J., Fitzgibbon, S. P., Lewis, T., Clark, C. R., Loveless, S., ... & Willoughby, J. O. (2007). Scalp electrical recording during paralysis: quantitative evidence that EEG frequencies above 20Hz are contaminated by EMG. Clinical Neurophysiology, 118(8), 1877-1888.

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

Wickens, C. D. (2008). Multiple resources and mental workload. Human Factors: The Journal of the Human Factors and Ergonomics Society, 50(3), 449-455.

Wiener, E. L. (1989). The human factors of advanced technology ("glass cockpit") transport aircraft (Contractor Report 177528). Moffett Field, CA: NASA Ames Research Center.

Wiener, E. L., & Curry, R. E. (1980). Flight-deck automation: Promises and problems. Ergonomics, 23(10), 995-1011.

Wientjes, C. J. (1992). Respiration in psychophysiology: methods and applications. Biological Psychology, 34(2), 179-203.

Wiesenthal, D. L., Hennessy, D. A., & Totten, B. (2000). The Influence of Music on Driver Stress. Journal of applied social psychology, 30(8), 1709-1719.

Wiesenthal, D. L., Hennessy, D. A., & Totten, B. (2003). The influence of music on mild driver aggression. Transportation research part F: traffic psychology and behaviour, 6(2), 125-134.

Wilde, G. J. S. (1982). The theory of RHT: implications for safety and health. Risk Analysis, 2, 209-226.

Wilson, G. F., & Russell, C. A. (2003). Operator functional state classification using multiple psychophysiological features in an air traffic control task. Human Factors: The Journal of the Human Factors and Ergonomics Society, 45(3), 381-389.

Wilson, G. F., & Russell, C. A. (2007). Performance enhancement in an uninhabited air vehicle task using psychophysiologically determined adaptive aiding. Human Factors: The Journal of the Human Factors and Ergonomics Society, 49(6), 1005-1018.

World Health Organization (2009). Global status report on road safety: time for action. Geneva: World Health Organization.

Zander, T. O., & Kothe, C. (2011). Towards passive brain-computer interfaces: applying brain-computer interface technology to human-machine systems in general. Journal of Neural Engineering, 8, 025005.

Zander T. O., Kothe C., Jatzev S., & Gaertner M. (2010). Enhancing human-computer interaction with input from active and passive brain-computer interfaces. In D. S. Tan & A. Nijholt (Eds.), The Human in Brain-Computer Interfaces (pp. 181-199), Berlin: Springer.

Zhao, C., Zhao, M., Liu, J., & Zheng, C. (2012). Electroencephalogram and electrocardiograph assessment of mental fatigue in a driving simulator. Accident Analysis & Prevention, 45, 83-90.

Zijlstra, F. R. H. (1993). Efficiency in work behaviour: A design approach for modern tools. Doctoral thesis. Delft, The Netherlands.