Lehrstuhl für Ergonomie

Measuring cognitive task load: An evaluation of the Detection Response Task and its implications for driver distraction assessment

Antonia Santina Conti-Kufner

Vollständiger Abdruck der von der promotionsführenden Einrichtung Munich Center for Technology in Society der Technischen Universität München zur Erlangung des akademischen Grades eines Doktors der Philosophie (Dr. phil.) genehmigten Dissertation.

Vorsitzende:		Priv.-Doz. Dr. Jörg Wernecke
Prüfer der Dissertation:	1.	Prof. Dr. Klaus Bengler
	2.	Prof. Dr. Josef F. Krems
	3.	Prof. Dr. Mark Vollrath

Die Dissertation wurde am 16.01.2017 bei der Technischen Universität München eingereicht und durch die promotionsführenden Einrichtung Munich Center for Technology in Society am 28.05.2017 angenommen.

What is modern technology? It too is a revealing. Only when we allow our attention to rest on this fundamental characteristic does that which is new in modern technology show itself to us.

Martin Heidegger, The Question Concerning Technology (1977, p.14)

Publications and supervised theses

The author of this thesis published as **first author** and **supervised** the following topic related papers and theses/student projects, respectively:

Papers

- Conti, A. S., & Bengler, K. (2014). Measuring driver distraction in dual-task settings. In *GfA - Frühjahrskongress.*

- Conti, A. S., Dlugosch, C., & Bengler, K. (2014). The effect of task set instruction on detection response task performance. In *de Waard, D., Brookhuis, K., Wiczorek, R., Di Nocera, F., Barham, P., Weikert, C., Kluge, A., Gerbino, W., and Toffetti, A., (Eds.), Proceedings of the Human Factors and Ergonomics Society Europe Chapter 2013 Annual Conference* (pp. 107-117).

- Conti, A.S., Dlugosch, C., Schwartz, F., & Bengler, K. (2013). Driving and speaking: revelations by the head-mounted detection response task. In *Proceedings of the Seventh International Driving Symposium on Human Factors in Driver Assessment, Training, and Vehicle Design* (pp. 362–368). Iowa City, USA: University of Iowa.

- Conti, A.S., Dlugosch, C., & Bengler, K. (2012). Detection response tasks: how do different settings compare? In *Proceedings of the 4th International Conference on Automotive User Interfaces and Interactive Vehicular Applications (AutomotiveUI '12)* (pp. 257–260). ACM Digital Library.

- Conti, A.S., Dlugosch, C., Vilimek, R., Keinath, A., & Bengler, K. (2012). An assessment of cognitive workload using detection response tasks. In *Advances in Human Aspects of Road and Rail Transportation* (pp. 735–743). CRC Press.

Supervised theses

- Laqua, A. (2016). Research on the relationship of simultaneous media use and cognitive failure with the multitasking capability in the vehicle. *Bachelor's thesis.*
- Winzer, O.M. (2016). Development of an index to evaluate the compatibility between secondary activities and the driving task. *Master's thesis.*
- Späth, M. (2016). The influence of interindividual differences on the result of the detection response task. *Master's thesis.*
- Späth, M. (2015). The effect of different vibration motors on detection response task metrics. *Bachelor's thesis.*
- Winzer, O.M. (2015). An investigation of the effects on cognitive workload and attention after distraction. *Semester thesis.*
- Huihui, D. (2015). Mobile detection task and auditory n-back task: a comparison of user-paced and system-paced task variants. *Bachelor's thesis.*
- Kellner, M. (2015). Power analysis to compare different variants of a cognitive workload measurement method. *Bachelor's thesis.*
- Baur, C. (2014). History and current status of IVIS. *Bachelor's thesis.*
- Späth, M. (2014). Detection response tasks: a comparison of placement, modalities, and systems. *Bachelor's thesis.*

Supervised experimental projects

- Dodwell, G. (2015). Comparative analysis of the detection response task and object event detections while driving – can the DRT be utilized to predict OED performance? *Neuro-Cognitive Psychology (NCP) semester break thesis.*

- Döllinger, N. (2014). Detection response task: evaluation of two different response types. *Ergonomics practical seminar thesis.*
- Helfer, B. (2012). The effect of task instruction on task performance. *Neuro-Cognitive Psychology (NCP) semester break thesis.*

Dedicated to those who always believed

Acknowledgements

I would like to thank my advisor, Klaus Bengler, for introducing me to ergonomics. I appreciate the time spent discussing my topic and guidance from day one both academically and personally. A special thanks goes to Carsten Dlugosch with whom I have worked on many experiments together and whose technical wisdom has done nothing less than impress and inspire me. Similarly, I would also like to thank Michael Krause, for his technical initiatives regarding the DRT and in other driver distraction studies we've performed together over the years. Thank you, Armin Eichinger, for providing me with continuous support and mentoring, especially for teaching me to relax in the face of experimental/publication stress. A big thank you to Fernando Cross Villasana and Bastiaan Petermeijer for their time and helpful feedback during the preparation of this thesis. For their statistical insight and advice, I thank Moritz Körber and Sean Seaman, especially the latter for our fruitful discussions early on regarding power analysis. I would also like to thank the members of the WG8 ISO group whom I had the honor of working with and learning from during the standardization process of the DRT. A special thank you to Johan Engström, who saved me from drowning in the cold waters of attention literature by suggesting key works. A warm thank you to my loving parents, Lee Ann and Anthony Conti, for educating me on the important things in life and providing me with the means to realize and accomplish my goals. Thank you to my brother, Christian Conti, for keeping me modest. Finally and most of all, I thank my husband, Maximilian Kufner, for his unending patience, support, and mathematical advice–you are truly my light in the darkness.

Abstract

The Detection Response Task (DRT) is an applied measure of the attentive effects of cognitive load and is evaluated in this thesis in terms of sensitivity, reliability, and validity. A review of empirical literature introduces the reader to concepts and topics relevant to understanding the DRT and its use in the field of driver distraction. Two empirical studies are presented. In the first experiment, the sensitivity of three DRT variants (viz., head-mounted [HDRT], remote [RDRT], and tactile [TDRT]) is tested. As DRTs are performed together with other tasks of interest, Performance Operating Characteristics provide insight regarding how the performance of tasks changes as a result of being performed together. Additionally, a power analysis is performed to provide an estimation of the strength of each of the DRTs in terms of required sample sizes. In the second experiment, electroencephalogram (EEG) is used to validate the DRT as a measure of the brain state of being cognitively loaded. Results support the sensitivity of the DRTs to variations in task load especially for artificial, cognitive tasks, as evident by higher RTs for more demanding tasks. The HDRT resulted as minimally obtrusive and affected concurrent task performance of other tasks the least. Additionally, the power analysis revealed the HDRT required the least amount of participants. Reliability of the DRT was also supported as the findings reported in Experiment I were able to be replicated in Experiment II. Validity of the DRT as a measure of a cognitively loaded state was not supported and a dissociation of the EEG and DRT measures is reported and discussed. Validity of the DRT in terms of its relevance on-the-road is discussed and proposed as a conditional relationship rather than a direct one. Specifically, the DRT measures how a person's attention and reaction to signals change as a result of being performed together with a task of interest, from which distraction potential may be inferred.

Abstrakt

Der Detection Response Task (DRT) ist ein Messinstrument zur Erfassung des Einflusses kognitiver Belastung auf die menschliche Aufmerksamkeit. In dieser Doktorarbeit wird die Eignung des DRT als Messinstrument anhand der Sensitivität, Reliabilität, und Validität untersucht. Eine Zusammenfassung der Literatur bezüglich bisheriger empirischer Studien führt den Leser zu Beginn in relevante Konzepte und Themen ein, um den DRT und dessen Verwendung im Kontext der Fahrerablenkung zu setzen. Es werden zwei empirische Studien präsentiert: in der ersten Studie wird die Sensitivität von drei Varianten des DRTs (head-mounted [HDRT], remote [RDRT], tactile [TDRT]) untersucht. Da der DRT immer zusammen mit einer anderen Aufgabe ausgeführt wird, kann mit Hilfe einer Performance-Operating-Characteristic-Analyse gezeigt werden, wie sich die Leistung in den Aufgaben infolge der gleichzeitigen Ausführung verändert. Darüber hinaus wird eine Teststärkenanalyse durchgeführt, um die Sensitivität der drei DRT-Varianten im Verhältnis zum Strichprobenumfang zu bestimmen. In der zweiten Studie wird der DRT, als Messmethode des kognitiven Belastungszustands mit Hilfe eines Elektroenzephalogramms (EEG) validiert. In den Ergebnissen zeigte sich, dass der DRT in Bezug auf unterschiedliche Belastungszustände, insbesondere bei künstlichen und kognitiven Aufgaben, sensitiv ist, was durch eine höhere Reaktionszeit bei schwierigen Aufgaben nachgewiesen wurde. Der HDRT erwies sich als am wenigsten störend für die Versuchspersonen und hat die anderen, gleichzeitig ablaufenden Aufgaben am wenigsten beeinflusst. Darüber hinaus hat die Teststärkenanalyse gezeigt, dass der HDRT den geringsten Strichprobenumfang zur Entdeckung eines Effekts benötigt. Die Reliabilität des DRTs konnte ebenfalls nachgewiesen werden, da sich die Ergebnisse des ersten Versuchs im zweiten Versuch reproduzieren ließen. Die Validität des DRT

als Methode zur Messung des kognitiv belasteten Zustands konnte nicht nachgewiesen werden. Darüber hinaus zeigte sich eine, in der Arbeit diskutierte, Diskrepanz zwischen den Messwerten des EEG und des DRT. Die Validität des DRTs als Messmethode im Realverkehr steht weiterhin zur Diskussion. Der DRT ist besonders sensitiv darauf, wie sich die Aufmerksamkeit und Reaktion einer Person auf ein Signal ändert, wenn gleichzeitig weitere Aufgaben bearbeitet werden. Aus den Ergebnissen des DRTs kann das Potenzial der Fahrerablenkung gefolgert werden.

Contents

Nomenclature and definitions

CC	Control command task; an auditory, working memory task requiring verbal input
Cognitive task	A task that does not require much or any physical (ex. visual, manual) interaction/manipulation to perform or achieve a specific goal
Cognitive task load	Characteristic of a task to be performed; demand placed on a user's mental resources or cognitive control by a task (see also ISO 17488:2016; terms and definitions section)
Counting task	An auditory, working memory task requiring mental arithmetic and verbal input
Distraction potential	Term used as a qualifier as the additional performance of a competing, concrete activity does not automatically mean that safety critical information is completely ignored; rather, that it is more probable under such circumstances
DRT	Detection Response Task (general term for the method)
Dynamic condition/scenario	Condition (also referred to as a scenario) performed with a driving or simulated driving task
FM	Frontal medial

HDBE	Tested condition; HDRT + simulated driving task + n-back easy
HDBH	Tested condition; HDRT + simulated driving task + n-back hard
HDN	Tested condition; HDRT + simulated driving task + no additional task
HDRT	Head-mounted Detection Response Task; DRT variant
HDVE	Tested condition; HDRT + simulated driving task + SuRT easy
HDVH	Tested condition; HDRT + simulated driving task + SuRT hard
HNBE	Tested condition; HDRT + n-back easy
HNBH	Tested condition; HDRT + n-back hard
HNN	Tested condition; HDRT baseline
HNVE	Tested condition; HDRT + SuRT easy
HNVH	Tested condition; HDRT + SuRT hard
HR	Hit rate
ISI	Inter-stimulus interval; see OnOn and OffOn
LCT	Lane Change Test
Load	Task difficulty characteristic; independent of the human operator and cannot be measured directly
Mental effort	Mental work required to perform at a specified level as per Brookhuis and de Waard (2001)
MR	Miss rate
N-back	An auditory, working memory task

OffOn	Inter-stimulus interval defined as signal offset to onset
OnOn	Inter-stimulus interval defined as signal onset to onset
Perception	The selection, organization and interpretation of sensory information into mental representations (Huffman, 2003, p. 128)
PSDs	Power spectral densities
RDBE	Tested condition; RDRT + simulated driving task + n-back easy
RDBH	Tested condition; RDRT + simulated driving task + n-back hard
RDN	Tested condition; RDRT + simulated driving task + no additional task
RDRT	Remote Detection Response Task; DRT variant
RDVE	Tested condition; RDRT + simulated driving task + SuRT easy
RDVH	Tested condition; RDRT + simulated driving task + SuRT hard
RMSE	Root-mean-square error (of speed)
RNBE	Tested condition; RDRT + n-back easy
RNBH	Tested condition; RDRT + n-back hard
RNN	Tested condition; RDRT baseline
RNVE	Tested condition; RDRT + SuRT easy
RNVH	Tested condition; RDRT + SuRT hard
RSM	Right sensory motor
RTs	Reaction times
SDLP	Standard deviation of lane position

Sensation	Receiving, translating and transmitting sensory information to the brain (Huffman, 2003, p. 128)
Sentences task	A realistic auditory, working memory task requiring verbal input and additionally required sentence comprehension
ST	Secondary tasks; tasks performed in addition to the DRT and or driving task. These tasks are used to induce different levels and types of demand
Static condition/scenario	Condition (also referred to as a scenario) performed without a driving or simulated driving task
SuRT	The surrogate reference task: a visual search task requiring manual input
Task difficulty	Level of complexity of a task (i.e., easy or hard). Related to the degree of task demand placed on a human performer; easy tasks are assumed to be less demanding than harder tasks
Task load	Characteristic of a task to be performed; level of demand or difficulty placed on the user by a task
TDBE	Tested condition; TDRT + simulated driving task + n-back easy
TDBH	Tested condition; TDRT + simulated driving task + n-back hard
TDN	Tested condition; TDRT + simulated driving task + no additional task
TDRT	Tactile Detection Response Task; DRT variant
TDVE	Tested condition; TDRT + simulated driving task + SuRT easy
TDVH	Tested condition; TDRT + simulated driving task + SuRT hard

TNN	Tested condition; TDRT baseline

Chapter 1

Introduction

The Cartesian notion of *thinking therefore being* explicates humans as thinking, reflective animals. Through our own experiences, we understand this and can recall explicit moments in which ideas or objects vividly captured our thoughts. Additionally, we have experienced moments of idle thinking or day dreaming of nothing in particular, just as we have moments of intense concentration and or mental manipulation to solve a task. These examples describe just a few ways in which we experience cognition. Human cognition as related to technology can be observed manifoldly. One example is the increasing *intelligence* of vehicles, machines, computers, devices, etc., implemented with the capacities to process, interpret and understand information in an approximately human-like fashion. As technology increasingly *cognates* and becomes more intelligent, the way humans interact with this technology also changes. Specific to the automotive domain, the beginning stages of this shift to the cognitive can be observed in in-vehicle technology trends where physical buttons and menus are replaced by intuitive voice command and speech control. A gift and a curse, although in-vehicle device functionality is thus no longer limited to that provided by physical buttons, whether, how and to which degree this new interaction form distracts the driver are not yet completely understood. This type of technology that *cognitively* or *mentally* (used interchangeably in this thesis) interacts with the human-user, often bypassing the need for physical interaction, and can manifest on the road as "degraded object and event detection" (Tijerina, 2000, p. 2, see also Figure 2.6 on page 32).

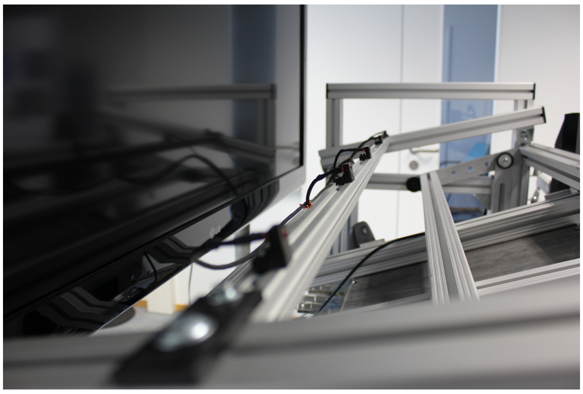
Many standardized measurement methods already exist but are not specifically sensitive to the type of attentive deficits associated with performing cognitive tasks. Such existing methods are: occlusion (standardized under: ISO 16673:2007), the Lane Change Test (LCT; standardized under: ISO 26022:2010), eye-tracking (standardized under: ISO 15007-1:2014 and ISO 15007-2:2014), and driving metrics (see SAE J2944 [2015]). As such, these measures can incompletely or erroneously measure the distraction potential of tasks that just affect one's ability to respond to an object or event. The aim of this thesis is to understand and investigate a measurement tool that attempts to quantify such performance impairments: The Detection Response Task (DRT), used to gauge task demand, especially cognitive demand, through a test person's ability to react to a presented signal.

1.1. Brief overview of the Detection Response Task (DRT)

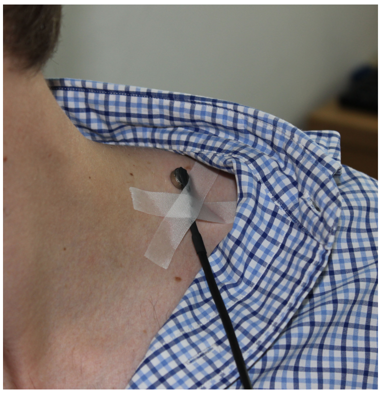
DRT variants can be found in Figure 1.1 on the next page. The DRT is a very simple method to use and implement in an experiment. With the DRT, signals are presented to a participant while he or she performs another task or other tasks. By pressing a button to indicate the DRT signal has been sensed and perceived, performance data are recorded and can be evaluated. Reaction times (RTs) and accuracy measures, such as hit or miss rates (HR, MR, respectively), are typically used for evaluations and comparisons of different devices, tasks (Olsson & Burns, 2000), and or experimental settings (Bruyas & Dumont, 2013; Jahn, Oehme, Krems, & Gelau, 2005). Based on these performance metrics, inferences can be made about the demand and distraction potential associated with the tested condition. Although the DRT can be used in many ways, its original purpose was to be used in driving studies. It is supposed that a task under evaluation yielding long DRT RTs and or a low HR, is more cognitively demanding and has a greater potential to distract a driver than if the RTs were shorter with a higher HR. Currently, the RT is considered the more sensitive metric (see T. A. Ranney, Baldwin, Smith, Mazzae, & Pierce, 2014), however, the HR and MR also convey important information about a person's ability to respond in adverse conditions (see van der Horst & Martens, 2010). In the real-world, longer

RTs and more frequently missed signals could be analogous to a missed street sign, not breaking on time in response to a sudden event, or even failure to break. A further discussion about the DRT can be found in section 2.6 on page 37.

(a) Head-mounted DRT (HDRT)



(b) Remote DRT (RDRT)



(c) Tactile DRT (TDRT)

Figure 1.1. – Examples of DRTs as used in the experiments presented in this thesis. (a) was placed on participants' heads. (b) was located between the test person and the simulated driving task display. Although 5 LEDs are shown, the center LED was not active. (c) was placed on the left shoulder area of the test persons.

1.2. Structure and goals of this thesis

The sensitivity, reliability and validity of the DRT are considered in this thesis (concepts as per Burns, Bengler, and Weir 2010, p. 25; but see also O'Donell & Eggemeier, 1986 for an earlier account) and are considered important criteria in determining the quality of a measurement tool. Sensitivity is understood as the capability of the DRT to reflect differences in task load (definition adapted from O'Donell & Eggemeier, 1986, p. 42-3); reliability, as the ability to reproduce DRT findings; and finally, validity, as concerned with understanding the measure of the DRT and its applied significance. Additionally, other important measurement tool criteria identified by Burns et al. and O'Donell and Eggemeier such as diagnostic and objectivity, will be discussed in sections 2.6.2, 3.3.1, 4.4.2 and 4.4.3.[1] Diagnostic is understood here as the capability to discriminate the demand-type placed on test persons (O'Donell & Eggemeier, 1986, p. 42-3) and to "pinpoint the locus of the overload" (O'Donell & Eggemeier, 1986, p. 42-6). In Chapter 2, the human operator is considered and a review of relevant literature on reaction times, attention, sensation and perception, driving and driver distraction is presented. The concepts covered in each of these sections provide a foundation upon which the DRT as a measurement tool may be conceived. The chapter ends with a comprehensive presentation of the DRT, explicating its current position in applied research. The first experiment is presented in Chapter 3 and describes an evaluation of the sensitivity of DRT variants to different levels and types of task load. In the second experiment presented in Chapter 4, an electrophysiological assessment of the DRT is reported in an attempt to validate the DRT as a measure of the cognitive state of being loaded. This assessment was implemented through a block-design using electrophysiological correlates of attentive processing and workload (alpha and theta, respectively). Additionally, metric reliability was evaluated in the second experiment. To summarize and conclude based on the literature and presented experiments, a general discussion is presented in Chapter 5.

---

[1] Method practicality, which is another criteria identified, was not within the scope of this thesis.

Chapter 2

Theoretical and empirical foundation

The aim of this chapter is to provide the reader with an understanding of DRT-related factors from historical, physiological, psychological, and applied perspectives. First, a brief history of using signal detection metrics, especially RTs, to measure human performance and selective attention will be presented. This will be followed by a presentation of different concepts of human attention, which is fundamental to the perception of stimuli as well as task performance in general. Various models of attention will also be described, detailing factors that can affect the way and how humans select, perceive, process and react to signals such as those presented in the DRT. In preparation for Chapters 3 and 4, which present experiments where visual and tactile DRTs were implemented, visual and tactile sensation and perception, as well as their relation to attention, are reviewed in the current chapter. The concept of cognitive workload is also discussed and its meaning is exemplified through its measure. In line with the use of the DRT in applied driving-related settings, literature on driving as a complex task as well as driver distraction, including existing and desired measurement methods, is presented. Here, the case is made for the need of a measurement tool, sensitive to tasks that might not have any obvious, directly observable distraction characteristics. Finally, the DRT method is reviewed in terms of its inception, previous findings, use, implications and on-the-road relevance.

2.1. The use of signal detection metrics to gauge human performance

Using signal detection metrics such RTs, HRs and MRs, to evaluate human performance has a long history dating back to the late 1700s. Around this time, it was believed that human processing and biological impulses were instantly communicated, "greater than the speed of light" (as per Johannes Müller's "Handbuch" as presented in Fuchs & Milar, 2003, p. 3); a theory which began to fade with developing research and advanced knowledge (Bolles, 1993; Fuchs & Milar, 2003). According to Bolles (1993), a primary turning point was an analysis by Friedrich Wilhelm Bessel of timing errors found in the observations of astronomers. In his analysis, Bessel found that different observers had different observation timing offsets and suggested to remedy such timing issues through calculating each individual's error constant, known as a "personal equation" (Bolles, 1993, p. 99; see also Proctor & Vu, 2003). A bit later on around the mid-1800s, Herman von Helmholtz performed studies showing signal transmission within the body was associated with a measurable time lag, i.e., a RT, but that sources of measure inaccuracy were present as "part of the measured time depends on mental processes" (Helmholtz, 1867, p. 228 as reported in Proctor & Vu, 2003, p. 295). The importance of using time measures to infer mental processes continued to grow and in terms of its relevance, Jastrow (1890) stated very early on:

> The study of the time-relations of mental phenomena is important from several points of view: it serves as an index of mental complexity, giving the sanction of objective demonstration to the results of subjective observation; it indicates a mode of analysis of the simpler mental acts, as well as the relation of these laboratory products to the processes of daily life; it demonstrates the close inter-relation of psychological with physiological facts, an analysis of the former being indispensable to the right comprehension of the latter; it suggests means of lightening and shortening mental operations, and thus offers a mode of improving educational

> methods; and it promises in various directions to deepen and widen our knowledge of those processes by the complication and elaboration of which our mental life is so wonderfully built up. (p. 99)

Although the use of RTs to measure physiological effects gained momentum over time, the pioneer work of two major scientists: Franciscus Cornelis Donders (1818-1889) and Wilhelm Wundt (1832-1920), contributed to its prominence in psychological studies.

Donders recognized early on that RTs could be used to measure the speed of mental processes. F. C. Donders (1969)[1] reported experimental findings (primarily from a doctoral student of his, De Jaager) where RTs for simple reaction tasks were performed more quickly than more complex tasks. One reported experiment detailed an investigation where participants were exposed to a stimulus delivered to the foot area and were to respond to the stimulus with the ipsilateral hand. In one condition, participants knew which side the stimulus was going to be delivered on and in a second condition, knowledge of which side the stimulus was to be presented was not had. Due to the increased complexity of task, the latter condition yielded longer RTs than the former. Donders concluded that the increase in RT was due to "the decision in a choice and an action of the will in response to that decision" (F. C. Donders, 1969, p. 419). Other experiments reported by F. C. Donders (1969) detailed differences in RTs to sound stimuli, for example, between a simple reaction task (responding to the presence of a stimulus; known as the "a-reaction"), a choice reaction task (discriminating stimuli that require distinct responses; known as the "b-reaction"), and a go/no-go task (responding to one stimulus [go] and ignoring others [no-go]; known as the "c-reaction") (Proctor & Vu, 2003; Robinson, 2001). Since each task type differed in terms of required mental processes, the differences observed in the RTs associated with each task type were directly related to the differences in mental processes. Thus, the basis for what is now referred to as the "subtraction method" was formed: time delays associated with simple tasks can be subtracted from the delays of more complex tasks, permitting the measure of the speed of the additional processes present in the complexer task and absent in the simpler one.

---

[1] Referenced version is the translated version. The original article was published in 1868.

Wundt is considered one of the founders of psychology as a field of study. Wundt's work on RTs used a technique referred to as self-observation, where test persons were trained to become aware of their psychological processes (Robinson, 2001, p. 168). Robinson (2001) reports that Wundt believed "mental reaction" (p. 169) to occur according to five stages:

> (1) sensation, the movement of the nerve impulse from the sense organ into the brain;
>
> (2) perception, the entry of the signal into the field of consciousness (*Blickfeld des Bewußtseins*);
>
> (3) apperception, the entry of the signal into the focus of attention (*Blickpunkt des Aufmerksamkeits*);
>
> (4) act of will, in which the appropriate response signal is released in the brain;
>
> (5) response movement, or more precisely, the movement of the response signal from the brain to where it initiates muscular movement. (p. 169)

Based on these stages, Wundt suggested that Donders method could be improved to measure stimuli discrimination through the measurement of a "d-reaction" where participants would perform an "a-reaction" task and respond only upon having identified the stimulus (Proctor & Vu, 2003; Robinson, 2001). The "d-reaction" was, however, highly criticized, especially by a student, James McKeen Cattell (1860-1944) (Robinson, 2001), and did not gain much popularity (Proctor & Vu, 2003). In 1887, the results of a student of Wundt's, Ludwig Lange (1863-1936), were published where RT differences could be explained by a participant attending to either the stimulus or a reaction (Robinson, 2001); thus opening "...up a line of reaction-time research on attention" (Robinson, 2001, p. 176).

2.1.1. Reaction time and accuracy rates

For a long time after the work of RT research pioneers, researchers maintained the notion that processing times were an additive sum of "separate times" (Luce, 1986, p. 96). However, Luce (1986) proposed understanding RTs as the sum of two component latencies: a "residual latency"

(p. 97) and "decision latency" (p. 96). Residual latency (R) is related to time delays associated with "transduction, transit, and motor times" (Luce, 1986, p. 97) and decision latency (D) is the time required for processing and reaction activation. Luce (1986, p. 97) describes this simple relationship as a formula for the observed RT (T) as follows:

$$T = R + D \tag{2.1}$$

Luce also acknowledged that RTs were able to be influenced by many different variables. In fact, in Chapter 2 of his book, Luce (1986, see pp. 49-51) advises caution with regard to using RTs in experimental settings. First, experimentally generating signals to be responded to does not exactly represent real world reactions to events (also discussed in section 2.6.4 of this thesis). Events that require signal detection in the real world come "haphazardly" (Luce, 1986, p. 50), different than those signals presented in an experiment. Second, other factors, such as response modality, can obfuscate RT and, therefore, the conclusions made based on the performance metric (e.g., finger versus foot response button/pedal). Luce (1986) also stated that participant-based variables that contribute to the general state of the test person, can also have an effect on the performance metric (further discussed here in section 2.6.2 on page 42). Boff and Lincoln (1988a), for example, indicated motivation and fatigue as inversely and directly related, respectively, to RT. Additionally, signal properties can also have an effect on RTs. For example, different mean RTs are to be expected depending on signal modality. Generally, under optimal conditions, RTs to visual signals occur around 150 ms post-onset and to tactile signals, around 110-120 ms (Boff & Lincoln, 1988c, p. 1842). Strong visual (Boff & Lincoln, 1988d, 1988e; Luce, 1986) or tactile (Conti, Krause, Späth, & Bengler, 2015) signals, are associated with quicker RTs than those produced by weaker signals. Strong signals also have a higher probability of being detected (Luce, 1986), which affects accuracy "a measure of the quality of a behavior" (Gawron, 2008, p. 14). When speedful responses are required, a possible trade-off between speed and accuracy may be observed, causing errors to rise when responses are quickly given, due to their speed. In order to avoid this trade-off in human behavioral studies, participants can be instructed to limit their error and attempt to control their behavior (see Heitz, 2014; Luce, 1986).

2.1.2. Reaction times used to investigate selective attention

In the 1900s, RTs became a widely used metric for investigations of *selective attention*, which can be understood as the selection and selective processing of stimuli (see Johnston & Dark, 1986 for a review of selective attention studies). Stroop (1935), for example, discovered what is known today as the "Stroop Effect", where naming the color of a spelled out color-name increases RT relative to reading the spelled out color-name despite its colored appearance. The implication of this study was that conflicting information requires additional processing time and the more dominant, practiced task (reading vs. color naming) would take precedence in terms of performance ease and speed. Ninio and Kahneman (1974) used RTs to investigate focused and divided attention. Participants were either exposed to two concurrent audio streams and instructed to either attend one (i.e., a focused attention task), both concurrently (i.e., a divided attention task), or they were exposed to one audio steam (i.e., single task: a focused attention task with only one audio stream). The authors reported that divided attention task performance was more erroneous and yielded higher RTs than focused task performance; however, higher errors were found for the focused task relative to the single task, with no mean RT difference[2] between them. Among others, the authors concluded that performing under conditions where multiple[3] tasks require attention is more difficult than when only one task requires attention. LaBerge (1983) used RTs to investigate the breadth or size of attention using a probe technique. Two experiments were reported where participants were presented with either words or nonwords and were to attend to either a letter or the entire string. Probes were also presented. Participants were instructed to react only when a target probe appeared and to ignore non-target probes. LaBerge (1983) reported that mean RTs to target probes were fastest for those presented in the middle of strings. Additionally, the breadth of attention was shown as task dependent since participants tasked with attending to entire strings did not show a RT benefit for probes

---

[2] A difference was, however, reported for the *SD* of the mean RT.

[3] Specifically, two tasks in Ninio and Kahneman (1974).

presented in different positions, however, a RT benefit was observed for participants attending to letters. Here, LaBerge (1983) was able to show through RTs that attention was able to zoom in or out according to the demands of the performed task.

2.1.3. Reaction times in applied research

In applied research, especially since the mid-1900s, researchers have used RTs to determine quality of task performance. In such cases, experiments are usually arranged such that a task of interest (i.e., a primary task) is to be performed either alone or together with another task (referred to as a secondary task). An example of performance evaluation through RTs of a primary task can be found in the experiment reported by Shinar and Vogelzang (2013). In this experiment, Shinar and Vogelzang investigated the speed and accuracy at which traffic signs were able to be understood based on display type (symbols vs. text). They found that presenting participants with only traffic sign symbols required more comprehension time, and therefore higher RTs, than when a text display was shown. In addition to being a measure of task performance, such paradigms can also be interpreted as indicators of task difficulty (as seen in Shinar & Vogelzang, 2013) and "operator workload" (generally understood as the *busyness* associated with performing a task [cognitive workload is discussed in detail in section 2.4 on page 23]), among others.

A dual-task paradigm ensues when two tasks (i.e., a primary and secondary task) are concurrently performed. Similar to the single task condition discussed in the previous paragraph, dual-task paradigms can also be used to indicate task difficulty and operator workload, in addition to being used to quantify, for example, "...the degree to which two tasks interfere with each other" (Luck & Vecera, 2002, p. 246) or the degree to which task performance changes when performed together with another task. Also referred to as the "secondary task technique" (Gopher & Donchin, 1986; O'Donell & Eggemeier, 1986; Ogden, Levine, & Eisner, 1979; Wickens & Hollands, 2000), in such paradigms two tasks are concurrently performed by a participant according to one of two applications (as per Gopher & Donchin, 1986; Knowles, 1963; O'Donell & Eggemeier, 1986): (1) the *loading task paradigm*, a primary task is performed and considered

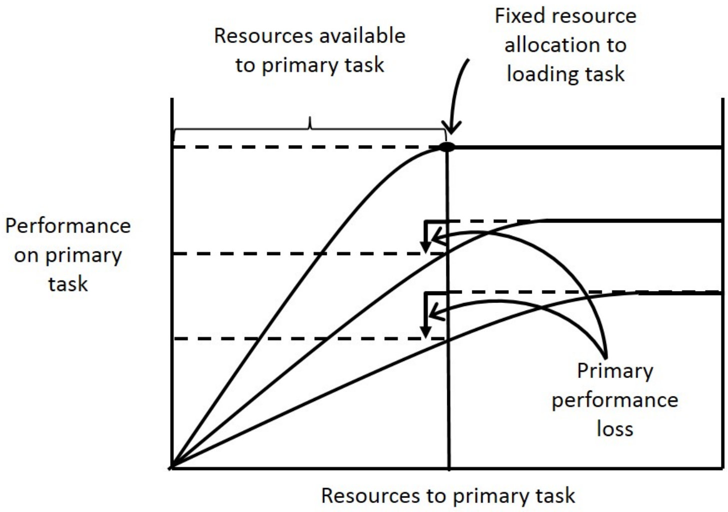


Figure 2.1. – Loading-task paradigm as illustrated by Wickens and Hollands (2000, p. 463). A primary task is performed together with another task–the loading task, which, depending on its difficulty, causes performance decrements of the primary task. The addition of a secondary task could be used to simulate aspects of a task, for example, that are not present in the experimental version of the primary task (Knowles, 1963, p. 156).

the "task of interest", and its performance is measured while the secondary task is performed (see Figure 2.1); (2) the *subsidiary task paradigm*, the primary task is performed at the same time as the secondary task, the performance of which is measured to quantify the residual performance or resource abilities *left over* from performing the primary task (see Figure 2.2 on the following page). An example of an experimental implementation of the loading paradigm using RTs was presented in Engström (2011, Paper III). In Engström (2011, Paper III), RTs of braking to a critical event on the road were used to investigate potential driving risks associated with performing cognitively loading tasks while driving. Specifically, all participants drove in simulated rural and urban areas and responded to a critical event on the road. The critical event was a vehicle traveling towards them in the opposite lane, unexpectedly turning left at an upcoming intersection and was presented to participants six times. Additionally, one participant sub-group performed an additional cognitive task (a cognitively loading, backwards-counting task) and another subgroup did not.[4] Engström (2011, Paper III) found no significant main effect of cognitive load on braking RTs, especially for initial reactions to the critical event, which as discussed by Engström (2011, Paper III), could be due to some reflexive, automatized, risk avoidance response.

---

[4] Within each sub-group different driving task instructions were also given: "half of the subjects were instructed to maintain a central lane position when driving on the rural road" (Engström, 2011, p. 8 of Paper III).

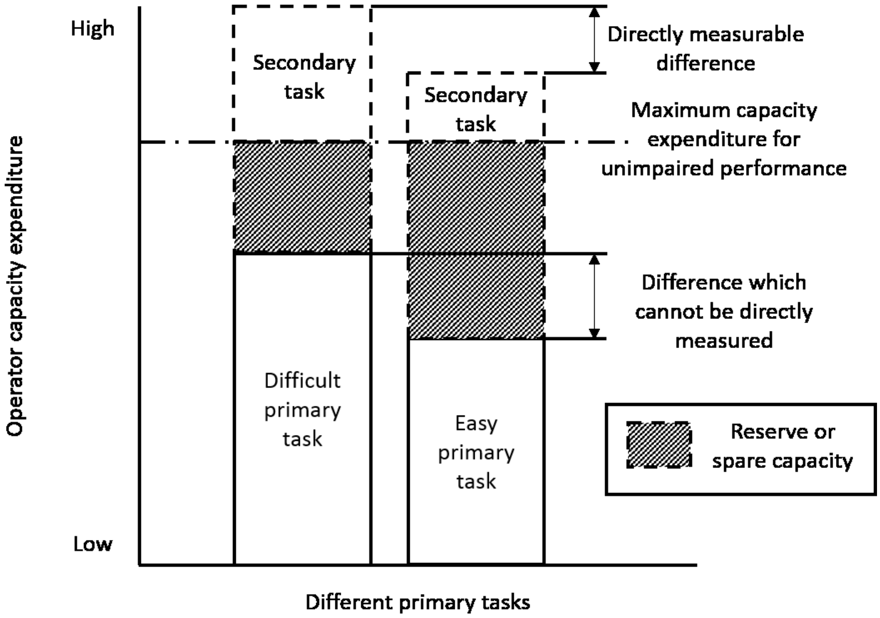


Figure 2.2. – Graphical depiction of how the subsidiary task paradigm measures operator reserve processing capacity as depicted by O'Donell and Eggemeier (1986, p. 42-25). A primary task is performed and its difficulty is reflected in the performance of the secondary task.

This study exemplifies the loading task paradigm as the researchers measured primary task performance under different conditions. An example of using RTs as a subsidiary task metric was reported by Patten, Kircher, Östlund, Nilsson, and Svenson (2006), where a detection task (i.e., the secondary task) was performed in addition to driving. In their experiment, the RTs of drivers with different levels of experience (high mileage, experienced drivers versus low mileage, inexperienced drivers) were compared for different levels of traffic environment complexity. Among other findings, Patten and colleagues reported that low mileage drivers had slower RTs than more experienced drivers in all traffic complexities. This finding suggests that due to their lack of experience, these drivers were less able to deal with the additional detection task despite traffic environment complexity. This study example exemplifies the subsidiary task paradigm as the researchers measured secondary task performance under different conditions.

2.2. Attention

Attention itself can be conceived of as serving a myriad of separate functions such as alerting and selecting (Posner & Boies, 1971); however, neither a central model of attention exists, nor is a definition of attention itself agreed upon. Describing the difficulty of defining an all encompassing term such as attention, Allport (1993) writes "there can be no simple theory of attention, any more than there can be a simple theory of thought" (p. 206). Attention, however, both in terms of being alert and selective processing, is necessary for signal detection and general task performance and as such needs to be understood. Mole (2012) suggests that attention is not only already known to us (as per the oft-cited James, 1890) but it can be considered "the phenomenon which explains the selective directedness of our mental lives" (p. 201). He furthermore distinguishes understanding attention in terms of two main approaches: *Broadbentian* (as per section 2.2.1), theorizing attention in terms of bottlenecks and a limited capacity, and *post-Broadbentian* (as per sections 2.2.2 and 2.2.3), where attention results from a contest between competing inputs (Mole, 2012). Early literature insisted that as humans are limited in capacity, attention was necessary to ensure only relevant stimuli were focused on and prioritized over less important signals, which would either be not selected (as per late selection theories, such as Deutsch & Deutsch, 1963), attenuated (such as A. M. Treisman, 1960), or not processed (as per early selection theories such as Broadbent, 1958; see Pashler, 1998b for a more detailed review on theories). In the following sections, attention will be reviewed in terms of theories and models relevant to the current thesis.

2.2.1. Applied Attention Theory and Multiple Resources

Possibly the most influential work in the area of applied attention has been published by Christopher D. Wickens, supported by a wealth of previous literature and theoretical bases (see Wickens & McCarley, 2008, pp. 130 - 132 for model history and rationale; as well as Boles, 2001 for an overview). The empirically based models proposed in Wickens (2002), Wickens, Hollands, Banbury, and Parasuraman (2013b), and Wickens and McCarley (2008) permit a general understanding of how attention operates and provides a basis on which performance prediction may be made. Wickens and McCarley (2008) identify five varieties of attention and their multifaceted functions: focused (the ability to concentrate on a task); selective (the ability to choose one object/task/stimulus over another); switched (the ability to move selective attention from one task to another); divided (the ability to parallel process); and sustained (the ability to endure and maintain attention) (pp. 1-3). In addition, Wickens and McCarley (2008) propose a simple and applied model of attention, according to which, attention is dualistic, functioning as a “filter” and the “fuel” behind information processing (both terms, p. 3). First, attention filters and selects incoming information from the external world based on any combination of top-down and or bottom-up factors or *pass* settings. Specifically, the bottom-up factors affecting the filter are considered the salience of a stimulus and effort in terms of cost (system, time, versus gain, etc.) and top-down factors are expectancy and value in terms of usefulness (Wickens, Hollands, Banbury, & Parasuraman, 2013a, pp. 50-53). The fuel (or attentional resources), which is limited, is allocated to further process incoming information (Wickens & McCarley, 2008).

Depending on the resources needed to perform in a multi-task situation, performance outcomes can be predicted by the dimensional multiple resources model (Wickens, 2002; Wickens et al., 2013b; Wickens & McCarley, 2008). According to this performance prediction model, when more than one task is required and if the resources of the input, processing or output of these tasks overlap (viz., intra-modal time sharing), performance decreases relative to single task performance or when no overlap occurs (viz., cross-modal time sharing) (Wickens et al., 2013b). In the model proposed in 2013, resources are divided into four separate and allocatable entities serving different stages of processing (viz., perception, cognition, and responding), perceptual modality (viz., auditory, visual or tactile), visual channel (viz., focal or ambient vision), and processing codes (viz., spatial or verbal), as well as response resources (viz., manual spatial and vocal verbal) (Wickens et al. 2013b, pp. 325-330; Wickens and McCarley 2008, pp. 132 - 137). Hence, sharing in terms of these distinct resources has a negative effect on performance.

2.2.2. Guided Activation Theory

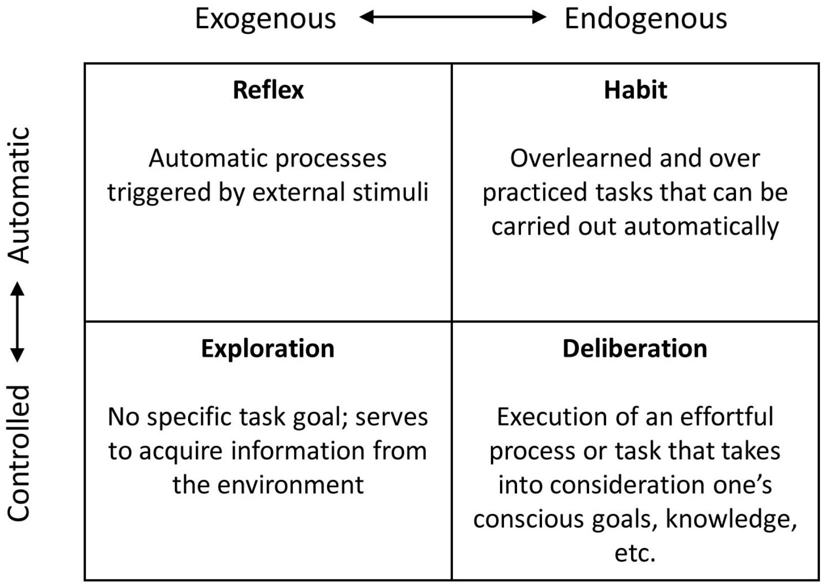
The Guided Activation Theory (GAT; J. D. Cohen, Dunbar, & McClelland, 1990) has its foundations in previous work on Parallel Distributed Processing (PDP; Rumelhart, Hinton, & McClelland, 1986) and Biased Competition (BC; Desimone & Duncan, 1995). The PDP model describes information processing as the result of a distributed series of activations, outputs and connections between processing units. In addition to these processing units, the general model necessitates an activation state, output functions, connectivity pattern, propagation, activation, and learning rules, as well as an environment where the system operates. These aforementioned factors determine which pattern or pathway is activated or selected and gives meaning to the processing units (Rumelhart et al., 1986, see pp. 45-55). Although Rumelhart et al. (1986) exemplifies the model through specific applications, the core model remains general enough to be specified by the researcher and or proposed elaborated model. The BC model explains attention as “an emergent property of many neural mechanisms working to resolve competition for visual processing and control of behavior” (Desimone & Duncan, 1995, p. 194). As such, the BC model proposes that selection of a stimulus or object for processing is the result of a competition across cerebral regions, biased by top-down (e.g. Chelazzi, Duncan, Miller, & Desimone, 1998) or bottom-up (e.g. Theeuwes, 1992) factors.

Combining and extending the PDP and BC models, GAT proposes that attention is a centralized, high-order, control mechanism originating in the prefrontal cortex (PFC), steering the activation of neuro-pathways relevant to and associated with the task being performed (Botvinick & Cohen, 2014; Miller & Cohen, 2001). Processing is considered a consequence of process type: automatic processes are known, over-learned tasks able to be processed without attention, and controlled processes, such as performing a novel task, require attention (J. D. Cohen et al., 1990). As a result, whereas automatic processes simply occur, controlled processes need to be attentively, effortfully performed to ensure performance. In terms of multiple task performance, GAT offers an alternative explanation for observed performance decrements when more than one task is performed concurrently; when two tasks activate the same network (viz., cross talk[5]), a control mechanism (viz., attention) is employed to pace (or “serialize” as per Botvinick & Cohen, 2014, p. 1269) processing as to not increase the demand on and overwhelm the already activated pathway (Botvinick & Cohen, 2014). Additionally, control allocation and task performance is dynamic: persons can “adaptively adjust” (Botvinick & Cohen, 2014, p. 1256) their control and improve performance (Botvinick & Cohen, 2014).

---

[5] Defined as “when two (or more) tasks make simultaneous demands on the same processing or representation apparatus” (Feng, Schwemmer, Gershman, & Cohen, 2014, Multitasking versus multiplexing section, para. 1).

2.2.3. Two-dimensional model of attention selection in driving



Automatic ⟷ Controlled	Exogenous ⟷	Endogenous
Automatic	**Reflex** Automatic processes triggered by external stimuli	**Habit** Overlearned and over practiced tasks that can be carried out automatically
Controlled	**Exploration** No specific task goal; serves to acquire information from the environment	**Deliberation** Execution of an effortful process or task that takes into consideration one's conscious goals, knowledge, etc.

Figure 2.3. – Visualizations of the two-dimensional model of attention selection as per Trick and Enns (2009) and Engström (2011). Identification of and description of the four possible modes of attention selection as proposed by Trick and Enns (2009, p. 65-67; figure on p. 66) and based on the adapted model presented by Engström (2011, p. 35) (the current figure is adapted: definitions added and task impairment omitted).

Trick and Enns (2009) suggest that attention selection in driving can be divided along two dimensions: controlled/automatic and exogenous/endogenous (bottom-up and top-down as per Engström, 2011 [see Figure 2.3]) processes, according to which four modes of selection may be categorized: reflex, habit, exploration, and deliberation. Reflexes and habits are those processes occurring automatically and are either triggered by external stimuli (exogenous) or overlearned and over-practiced (endogenous), respectively. Contrarily, the exploration and deliberation modes of attention selection are controlled processes, serving to acquire environmental information (exploration; exogenous) and to execute an effortful process (deliberation; endogenous). Accordingly, although automatic processes may occur in parallel without much effort, a controlled process requires effort and can interfere with other controlled processes (Trick & Enns, 2009). Additionally, controlled processes have also been found to be affected by additional cognitive load. In Engström (2011, Paper III), the ability of drivers to behave automatically (i.e., respond to threatening sudden events) was not affected by additional working memory load.

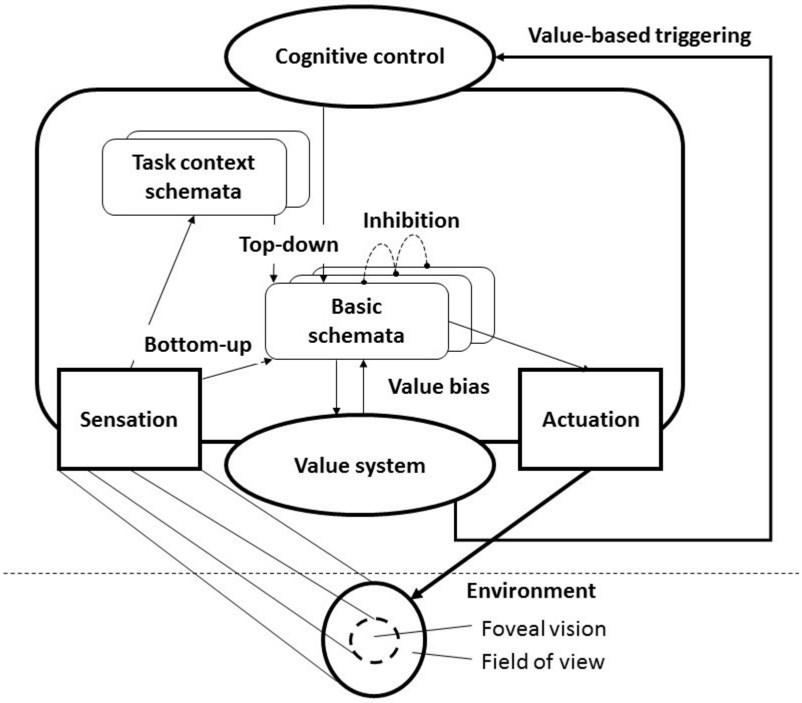


Figure 2.4. – Attention selection model proposed by Engström, Victor, and Markkula (2013, p. 34). "Arrows represent excitatory and dots inhibitory links" (Engström, Victor, & Markkula, 2013, p. 34). See text for an explanation of the model.

However, cognitively loaded drivers performed less well on controlled tasks, showing less adaptive braking behavior to on-the-road events presented overtime and shorter glances to oncoming vehicles (with a non-significant trend to look more often), than non-loaded counterparts Engström (2011, Paper III).

In 2013, Engström, Victor, and Markkula elaborated the model of Trick and Enns (2009). Central to this model proposed by Engström, Victor, and Markkula (2013) are *schemata*, which are "knowledge structures" or "units of action control" (both terms, p. 33) containing learned action/reaction sequences. In their conceptual model of attention selection (see Figure 2.4), attention is that which results from the process of schemata selection, "a state defined by a set of active schemata" (Engström, Victor, & Markkula, 2013, p. 35). Attention selection is understood as an "adaptive behavior" (Engström, Victor, & Markkula, 2013, p. 33), which maintains a balance between ultimate and process goals (e.g., arriving at a certain destination, safely) through schema or schemata selection (Engström, Victor, & Markkula, 2013). As in Trick and Enns (2009), the selection of schemata can occur automatically or controlled, the latter employed for novel situations without schema or to resolve a conflict between schemata. Additionally, schema selection is affected by a combination of bottom-up and top-down biases such as those transferred from the object or stimulus itself (viz., sensory input bias), behavioral relevance (viz., value bias), specific context setting (viz., contextual bias), and effortful directives (viz., cognitive control[6]) (Engström, Victor, & Markkula, 2013, p. 40). Similar to the GAT model described in section 2.2.2, schemata compete to be selected and the highest activation is *awarded* with selection.

According to Engström, Victor, and Markkula (2013), and in explanation of their the model presented in Figure 2.4, information is received and a proper reaction to a given situation occurs through the activation of schemata. Whereas task schemata relate generally to performing a task, basic schemata are subdivided into sensory-motoric and semantic schemata, which are involved in selecting an action to execute and in perceiving the situation, respectively. The selection of schemata can occur in both a bottom-up and top-down way. Inhibition between schemata demonstrates their competition for selection. A value system, which determines stimulus relevance, also acts as a bias for the selection of schema or schemata. This value system bias of schemata selection is established "either innate ('hardwired' from birth and selected through evolution) or learned through experience" (Engström, Victor, & Markkula, 2013, p. 37). Additionally, the value system can mobilize cognitive control, which also biases selection through intentionally increasing activation of a schema according to need. Cognitive control is fundamental to schema activation under circumstances where a weak but appropriate schema needs an added *push* to reach selection. Additionally, in the case of schema conflicts such as "[t]he lack of sufficiently strong habitual schema to match the current situation" or "[a] mismatch between habitually selected schemata and the current situation" (both from Engström, Victor, & Markkula, 2013, p. 38)[7], cognitive control is required to resolve attention selection. As a re-

---

[6] Based on Botvinick, Braver, Barch, Carter, and Cohen (2001), Anguera et al. (2013) states that "[c]ognitive control is defined by a set of neural processes that allow us to interact with our complex environment in a goal-directed manner" (p. 97). In the original work by Botvinick et al. (2001), cognitive control represents adaptability and is referred to as "[a] remarkable feature of the human cognitive system... to configure itself for the performance of specific tasks through appropriate adjustments in perceptual selection, response biasing, and the on-line maintenance of contextual information" (p. 624).

[7] Both excerpts were originally italicized and this emphasis has been omitted here.

sult, actions/reaction sequences are actuated based on the interplay of these aforementioned subcomponents.

In terms of dual-task interference, Engström, Victor, and Markkula (2013) consider three different possibilities that would cause multiple task performance interference: a physical reason for miss or response delay, cross talk of activated schemata, or a shared demand for control, referred to as peripheral, structural, and control interference (p. 44), respectively (pp. 44-45).

2.3. Visual and tactile sensation & perception

At the very basic level of signal detection theory, a signal needs to be sensed and perceived before any meaningful reaction can be made to it. In this section, literature on visual and tactile sensation and perception will be presented.[8]

2.3.1. Visual

The portion of the electromagnetic spectrum visible to humans, known as light, is between 380-780 nanometers (Boyce, 2006, p. 644). Vision occurs through a combination of eye anatomy, photoreceptors located in the human eye as well brain regions involved in vision, such as the visual cortex. Vision begins when lights enters the eye through the cornea and lens and reaches the retina. The retina is sensitive to light through two types of photoreceptors: rods, active in dim or dark settings, and cones, active in bright settings (Boyce, 2006; Mather, 2011). Cones also play a role in color vision as different cone subtypes are sensitive to different portions of the electromagnetic spectrum: specifically, long (red), middle (green) and short (blue) wavelengths (Boyce, 2006; Wässle, 2004). Located on the retina, the fovea is used in focused vision, having the highest concentration of cones and no rods. Rods are more active in peripheral vision and are highly concentrated on the retina away "from the fovea, reaching... maximum concentration around 20°" (Boyce, 2006, p. 651) from it. When light hits the retina, it is transduced into a series of chemical reactions. First, signals are transferred to the outer plexiform layer,

---

[8] Only modalities relevant to the experiments presented in Chapters 3 & 4, i.e., visual and tactile, are presented.

consisting of bipolar and horizontal cells, onto the inner plexiform layer, occupied by amacrine cells–inhibitory inter-neurons, and ganglion cells–which pass signals to the brain through the optic nerve (Wässle, 2004). As the "retina is organized such that increasing numbers of photoreceptors are connected to each optic nerve fiber as the deviation from the fovea increases" (Boyce, 2006, p. 652), humans are highly sensitive to the visual periphery, which is fundamental to detecting stimuli presented in this area (Boyce, 2006).

Visual perception and attention are related. It has been found that some peripherally presented cues capture attention more effectively than those centrally presented (Jonides, 1981). Jonides (1981) reported that the mean RTs of a target identification task were faster for peripherally cued targets than for those centrally cued. They also reported it more difficult for participants to ignore peripheral cues when instructed to do so, evident by a RT benefit for valid cues and cost for invalid cues, not observed for those centrally presented. Additionally, the mean RTs to peripherally cued targets were relatively unaffected by concurrent working memory load and expectation (Jonides, 1981). Not only is attention captured by peripheral stimuli, eye movements also seem to be captured subconsciously by peripheral stimuli (Theeuwes, Kramer, Hahn, & Irwin, 1998). In their experiment, Theeuwes et al. (1998) required participants to saccade to a peripherally presented, gray target circle and to identify the letter within it. In some trials, an additional task irrelevant, distractor circle appeared at the same time as the target. In these trials, participants were unaware of any changes in their eye movements, however, saccades often traveled first to the new object rather than to the target and increased target letter RTs. This effect, however, disappeared when the target was cued beforehand (Theeuwes et al., 1998). In sum, literature supports that peripheral stimuli can shift attention both covertly (as per Jonides, 1981) and overtly (as per Theeuwes et al., 1998).

2.3.2. Tactile

The sensation of vibration is mostly sensed by cutaneous or subcutaneous (Halata & Baumann, 2008) mechanoreceptors, which are "a class of sensory receptor cell that responds to mechanical distortion or deflection" (Mather, 2011, p. 22). Mechanoreception is mediated by four channels:

P (Pacinian), NP (non-Pacinian) I, NP II and NP III channels (Bolanowski, Gescheider, Verrillo, & Checkosky, 1988), together permitting humans to sense vibrations above .4 Hz (Bolanowski et al., 1988; Gescheider, Bolanowski, & Verrillo, 2004; also reviewed in Jones & Sarter, 2008). "Some mechanoreceptor types are rapidly adapting and respond at the onset and offset of stimulation, whereas others are slow adapting and respond throughout the time that a touch stimulus is present" (Proctor & Proctor, 2006, p. 75). From the skin, nerves carry the tactile signal to the spinal cord, eventually reaching the somatosensory cortex, located in the parietal lobe and involved in tactile processing (Hsiao & Yau, 2008; Purves et al., 2001). The sensation of tactile signals, and eventually also tactile perception, depends on the region on the body stimulated, the receptors in that area to receive the signal, as well as the cortical area dedicated to the stimulated portion of the body. These factors are related as areas of the body with many receptors also have larger cortical representations (Békésy, 1957; Marieb & Hoehn, 2016, p. 458). However, most areas of the body are optimally sensitive to vibrations between 150 - 300 Hz (Jones & Sarter, 2008, p. 91).

Tactile sensation and perception are very closely related to and affected by attention selection (Müller & Giabbiconi, 2008). Müller and Giabbiconi (2008) presented and discussed the idea that because the sensation of tactile signals necessitates close proximity to the human perceiver, they are different than visual and auditory signals. Additionally, simple tactile detection tasks have been found to result in faster human RTs than visual detection tasks (Ng & Chan, 2012). Arguably, the importance of being able to quickly process tactile signals could be evidence of an evolutionary strategy. Hanson, Whitaker, and Heron (2009) argued that tactile stimuli could be automatically processed because reacting to them is not affected by accompanying additional sensations. This argument was determined based on an experiment where Hanson et al. (2009) tested participants RTs to visual, auditory, and tactile stimuli under unimodal (only one modality to respond to) and dual-modal (two modalities used simultaneously) conditions. In the unimodal condition, no significant differences were reported between the RTs to each signal modality. However, in the dual-modality condition, whereas RTs to visual and auditory stimuli suffered, RTs to tactile stimuli were robust. Additionally, tactile signals are able to capture visual attention and facilitate responses to critical events. In a study reported in Ho and Spence (2008, Chapter 5), the effectivity of tactile signals to direct attention to driving relevant events was tested. The experimenters had participants perform two tasks simultaneously: (1) a visual task where target numbers were to be detected and distractor letters were to be ignored; (2) a simulated driving task where video clips of driving scenarios were presented with critical events (viz., either a vehicle behind the participant's virtual vehicle suddenly came too close or the participant's vehicle traveled too close to a vehicle in front of it) periodically required a reaction (viz., activating the brake or gas pedal) from the participant. Tactile signals were delivered through a vibration belt worn around the waist with motors placed on the stomach and back area. Those signals that accurately predicted the direction of the critical event, i.e., in front of or behind the participant, facilitated the actual response to the critical event. In sum, tactile signals are salient as well as effective in directing attention.

2.4. Cognitive workload

The concept of cognitive workload has grown in popularity in empirical and applied research since the 1980s (M. S. Young, Brookhuis, Wickens, & Hancock, 2015). Similar to attention, cognitive or mental workload has no singular, agreed upon conceptual or operational definition. According to Parasuraman and Caggiano (2002), "[m]ental workload refers to a composite brain state or set of states that mediates human performance of perceptual, cognitive, and motor tasks" (p. 17). Generally, mental workload can be understood as an individual and task based (also, top-down and bottom-up, respectively) amount of information processing capacity used to perform a task or task set (de Waard, 1996, p. 15). Many conceptualizations of cognitive workload stipulate human information processing capacity be limited, accounting for the human inability to concurrently perform an infinite amount of tasks (see Gopher & Donchin, 1986, p. 41-3). The concept of mental workload has also been used to refer to a more general "amount of cognitive resources involved in performing a task" (Sammer, 2001, p. 350). Many also use this general term to access answers to questions similar to those identified in M. S. Young et al. (2015):

> How busy is the operator? How complex are the tasks that the operator is required to perform? Can any additional tasks be handled above and beyond those that are already imposed? Will the operator be able to respond to unexpected events? How does the operator feel about the tasks being performed? How many people are needed to successfully carry out the task? (p. 1)

Effectively, the aforementioned questions attempt to access a similar fundamental matter: how does a task of interest affect a user's cognition, attention, and ability to perform? The value in investigating the aforementioned research questions is to understand the dynamic between task demand, cognitive supply, and potential performance breakdowns. Because of its conceptual broadness, cognitive workload is often referred to as a hypernym rather than directly addressed in terms of related factors/concepts such as cognitive task load and selective attention. In the following section, measures of cognitive workload will be discussed, which will provide a more comprehensive idea of what this concept indicates.

2.4.1. Understanding cognitive workload through its measurement

As task demand cannot be directly measured, measures of *cognitive workload* are often suggested as a way to quantify and or qualify task-related effects on human cognition, attention and performance. Cognitive workload is not associated with any one unanimous, specific metric and it is often operationalized on a case-by-case basis. Cognitive workload can manifest itself in many ways and can be assessed through 4 main techniques: primary task performance, secondary task performance, subjective and physiological measures (de Waard, 1996; O'Donell & Eggemeier, 1986; Wickens & Hollands, 2000). In order to have a more complete understanding of the workload associated with a particular situation, de Waard and Lewis-Evans (2014) suggest:

> Specifically, multiple measures should be taken, performance, self-reports, and if possible physiology, and, very importantly, these do not need to correlate, else the assessment of one type would suffice. Dissociation of measures gives a better view on what has happened during performance of a task, what strategies were applied, and whether the operator had to try hard to protect performance. (pp. 304-305)

Only when various measures are taken can a holistic understanding of cognitive workload as a state begin to be understood.

The first two techniques were already presented in section 2.1 in terms of RTs. Despite operationalization and metric used, there is a presumed direct relationship between cognitive workload and task performance. O'Donell and Eggemeier (1986, see p. 42-3), for example, proposed that the relationship between mental workload and performance is inverse after a person's capability has been exceeded until performance can no longer further decrease. When task load and workload are low, task performance is expected to be high and even slight increases in load will not cause performance detriments. However, as soon as load increases beyond a person's capability, workload also rises and task performance gradually suffers until a lower-limit is reached.

The aforementioned description of the relationship between workload and performance is not perfect: sometimes task performance does not change as a result of increasing workload (de Waard & Lewis-Evans, 2014). From this perspective, cognitive workload is related to *mental effort*. Mental effort and workload are not precisely the same as delineated by Brookhuis and de Waard (2001): "[e]ffort is a voluntary process under control by the operator, whereas mental workload is determined by the interaction of operator and task" (p. 322). Brookhuis and de Waard (2001) adds that "[a]s an alternative to exerting effort, the operator might decide to change the (sub)goals of the task" (p. 322). De Waard and Lewis-Evans (2014) gave the example of an operator increasing his or her effort in order to maintain a certain level of task performance; referred to as "performance protection" (Hockey, 1997 as referenced by de Waard & Lewis-Evans, 2014, p. 304). In accordance with this, task performance is not sensitive to the additional workload generated by, in this case, increasing effort to perform a task and therefore cannot solely and completely (de Waard & Lewis-Evans, 2014) estimate workload. Under such circumstances, subjective and or physiological measures might prove more sensitive to different aspects of workload. In terms of subjective measures, Brookhuis and de Waard (2001) distinguishes between "self-reports" and "(expert) judgments" (p. 325). Self-reports involve the operator reporting his or her experience during some event. Self-reports are often acquired through already validated questionnaires such as SWAT (subjective workload assessment technique), NASA-TLX (task load index; not specific to just mental processes), RSME (rating scale of mental effort) (see Brookhuis & de Waard, 2001) and DALI (relevant for driving related studies: driving activity load index; Pauzié, 2008). Expert judgments regarding task performance may also be made, which involves a performance quality evaluation by an expert or well-versed observer (see Brookhuis & de Waard, 2001, pp. 325-326).

Physiological measures record activity fluctuations in the human body. Although a more comprehensive documentation of the plethora of physiological measures of cognitive workload was published by de Waard (1996), frequency analyses of theta (traditionally: 4-7 Hz; Steriade, 2005, p. 31) and alpha (traditionally: 8-13 Hz; Steriade, 2005, p. 74)[9] through EEG are focused on due to their relevance in Chapter 4 of this thesis.

Accessing workload through EEG: Alpha and theta

EEG[10] provides one possible way of recording (electro)physiological measures of workload. EEG itself provides "a window into the human brain" as put by Sauseng and Klimesch (2008, p. 1002) in terms of mental state and activity level determination. Through the placement of electrodes on the scalp of test persons, the extracellular cortical electrical activity that occurs during a period of time or experimental condition is amplified and recorded (for a detailed explanation, see Speckmann & Elger, 2005). Activity is determined as the difference between the active (e.g., active cortical area) and reference (e.g., less active area) electrodes. Event-related potentials (ERPs) and power spectral densities (PSDs) can be derived from the EEG; which method is selected, depends primarily on the research question. ERPs are primarily used for mental chronometry, especially for those research questions aimed at identifying the speed and intensity of perceptual, cognitive, and motoric components underlying a certain behavior. Some consider ERPs as "high-tech substitutes for reaction-time measurements" (Luck, Woodman, & Vogel, 2000, p. 433). Alternatively, analyses involving PSDs can be observed in a block design or as event-

---

[9] Although frequency band ranges are reported, these ranges may vary per individual based on a number of factors (see Klimesch, 1999). Klimesch (1999) suggests to adjust these ranges to fit the individual test person for more accuracy.

[10] Parts of this general EEG summary were written by the author of this thesis and published in German (see Bubb et al., 2015, pp. 643-644).

related (see Wickens et al., 2013b, pp. 353-355 for an overview). In such analyses, through a Fourier transform (see Bruns, 2004, for a review and discussion of such signal analysis approaches), the EEG signal is decomposed into component frequency bands (viz., alpha, theta, beta, gamma [the latter two: 20-60 Hz; Steriade 2005, p. 31]). Block designs are typically implemented to evaluate brain state. Depending on the task demand and brain state of a test person, different wave frequencies can become more or less "dominant" (Teplan, 2002, p. 2) in the EEG signal (Teplan, 2002). "Additionally, certain EEG parameters are certainly potentially very useful as an index for mental workload. Frequency analysis of brain activity renders a picture of distribution of activity across frequency bands, serving as an indication for driver state with respect to [for example] alertness and vigilance in different driving conditions" (Brookhuis & de Waard, 2001, p. 329; in reference to Brookhuis, 1995).

Interpretations and relevance of wave frequencies have already been established in empirical literature. Tsang and Vidulich (2006) report on Parasuraman and Caggiano (2002) stating that "[s]pectral power[11] in the alpha band that arises in widespread cortical areas is inversely related to the attentional resources allocated to the task, whereas theta power recorded over the frontal cortex increases with increased task difficulty and higher memory load" (p. 256). Therefore, low alpha amplitude (also referred to as "desynchronized") is assumed to represent more attentional allocation than a high alpha amplitude (also, "synchronized"). For theta, synchronized theta is assumed to represent increased task load relative to desynchronized theta, representative of less task load. The following paragraphs will provide an review of select existing literature on theta and alpha as workload indices. It is useful to note, however, the majority of these studies use event-related settings rather than block-design.

Klimesch (1999) suggested that theta amplitude typically rises during task engagement and could reflect "unspecific factors such as e.g., attentional demands, task difficulty and cognitive load" (p. 187), as well as "working memory or episodic memory" (p. 185). Almahasneh, Chooi, Kamel, and Saeed (2014) used theta to infer driver cognitive distraction (see also Lin, Chen, Ko, & Wang, 2011). In their experiment, participants drove with and without an additional cogni-

---

[11] Amplitude squared (Bruns, 2004).

tive secondary task while EEG was recorded. In terms of theta, they reported that more difficult secondary tasks were associated with a higher theta level in frontal areas. Frontal-midline theta has also been reported as higher in conditions where a novel movement sequence is performed relative to a known sequence (Sauseng, Hoppe, Klimesch, Gerloff, & Hummel, 2007). Specifically, Sauseng et al. (2007) concluded that especially when a certain performance criteria is mandated (e.g., an instruction to perform as accurately as possible), cognitive resources are additionally recruited and allocated in order to meet task demands. This increase of cognition is reflected electrophysiologically as an increase of frontal-midline theta. Frontal-midline theta amplitude has also been found to increase with the demands of both visuo-verbal and visuo-spatial (Gevins, 1997) working memory tasks and tasks requiring cognitive control (Anguera et al., 2013; Cavanagh & Frank, 2014).

According to Klimesch (1999), alpha is inversely related to theta, reflecting "task demands and attentional processes" (p. 183) and "semantic memory processes" (p. 185). Generally, active brain areas show decreased alpha levels relative to non-active areas (Jensen & Mazaheri, 2010). Alpha amplitude is inversely related to theta and task difficulty, reflecting cognitive, perceptual and attentional processes (see Klimesch, 1999; Klimesch, Sauseng, & Hanslmayr, 2007; Sauseng & Klimesch, 2008, for reviews). In the study by Gevins (1997), mentioned in the previous paragraph, increasing task difficulty was associated with a decrease in alpha amplitude measured in parietocentral and occipitoparietal areas. Frey et al. (2016), for example, reported that the alpha amplitude of task-relevant sematosensory areas was suppressed prior to tactile signal presentation. This pre-stimulus decrease in alpha has been interpreted as reflecting high excitability (Weisz et al., 2014) or expectation (as per lower-2 alpha reported in Klimesch, 1999, pp. 183-185).

Lei (2011) and Lei and Roetting (2011) reported using both theta and alpha to investigate driver workload. Specifically, Lei and Roetting (2011) reported that frontal theta increased and parietal alpha decreased with increasing working memory load (manipulated through a visual n-back task). Additionally, Lei and Roetting (2011) also reported that driving demand manipulations where drivers either watched a driving scene or performed lane change maneuvers at

75 or 100 km/h were associated with an decrease in parietal alpha power for more demanding conditions, leaving frontal theta statistically unaffected.

2.5. Driving and driver distraction

In the previous sections, human attention, sensation, perception, and cognitive workload were presented. Beyond these topics, as this thesis focuses on an applied use of the aforementioned *human factors* in driving, it is useful to understand driving as a task, the distraction away from this task, as well as how these are operationalized and measured.

2.5.1. Driving as a concept

Nowadays, most persons know what driving is by having driven themselves and/or having been driven to some destination. Seen empirically, driving is a task that requires the simultaneous execution of multiple subtasks. According to Bubb (2015), driving can be divided into three main subcomponents: primary, secondary, and tertiary tasks. Primary tasks are those having the highest priority in terms of task goals (i.e., transportation) and timeliness. Specifically, these subtasks are identified as: navigation, operation, and stabilization of the vehicle. Secondary tasks within this model are those that support the priorities identified for the primary task. Examples of these would be signal indicator activation/deactivation and headlight activation in response to darkness. Tertiary tasks makeup the final category, consisting of comfort and communication relevant tasks such as adjusting temperature, seating, and radio controls, as well as managing telecommunications, Internet, etc. This last category of tasks indirectly contribute to driving in terms of how the driver and passengers feel within the vehicle and experience the ride. Extreme cold or heat, for example, has been found to worsen lateral vehicle control (Daanen, van De Vliert, & Huang, 2003).

Driving may also be conceived as a "control task in an unstable environment created by the driver's motion with respect to a defined track and stationary and moving objects" (Fuller, 2011, p. 13). According to the driver control theory by Fuller (2005; 2011), when a driver operates a ve-

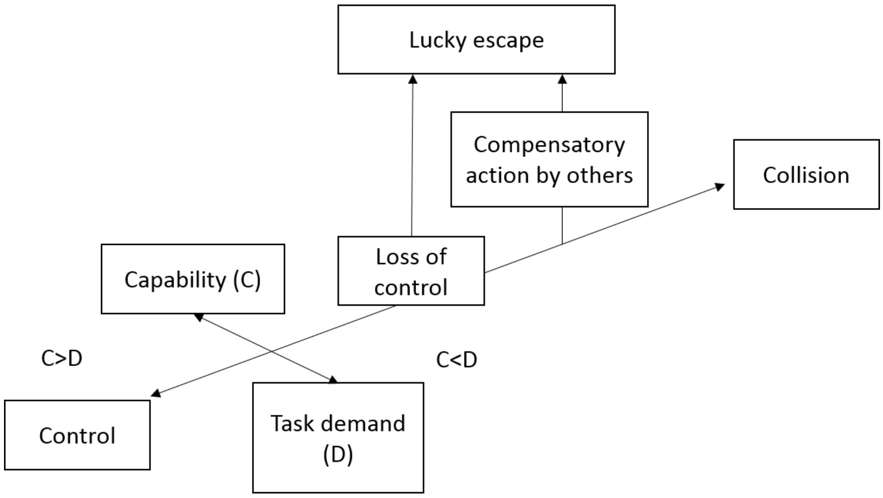


Figure 2.5. – The basic "task-capability interface" model from Fuller (2011, p. 14; adapted with formatting changes). See text for explanation.

hicle, constant and continuous comparisons are made between perception and behavior, which serves to direct the driver's actions and reactions. In order to ensure safety and ideally, the driver maintains a balance between the demands of the task being performed with his or her own capability. This is described in Fuller's "task difficulty homeostasis" model (2005; 2011), where drivers will take on or search for additional tasks to perform when the situation permits (e.g., changing the radio station on the highway with no traffic in sight) and cease to perform excessive tasks when the situation requires (e.g., ceasing to adjust the radio when traffic is present and a lane change is needed). According to Fuller's basic "task-capability interface" model (Fuller, 2011; for the complete model, see Fuller, 2005, p. 465) depicted in Figure 2.5, as long as this balance between demand and capability is maintained, and the capability either meets or exceeds the task demand, safe driving should ensue. However, as soon as the task demand exceeds capability, a loss of vehicle control occurs. Under such circumstances, a collision is imminent unless a critical situation is absent or other traffic participants compensate for this loss of control by changing their own behavior. Within Fuller's model, *driving demands* can be understood as vehicle controllability and the information acquisition ease, and, *capability*, as the ability of the driver to perform within his or her own physical limitations or that of the vehicle being operated, based on education and previous experience (both concepts as per Fuller, 2011, p. 15). "In principle, the greater that capability is, relative to task demand, the lower the difficulty of the task and vice versa" (Fuller, 2011, p. 14).

In line with Fuller's model, previous literature supports that drivers self-pace and balance the demands of a task(s) to meet their own capability (see Fuller, 2011; R. A. Young, 2015; as well as Figure 2.5 on the previous page). It has been seen, for example, that under normal circumstances, if drivers are given the choice to perform a secondary task at will, they will perform such tasks on easy to drive road segments where timing is not as critical as during more complex road conditions (as reported in Summala, 2007, see especially section 11.4). Additionally, drivers have also been found to adjust their performance of other tasks (e.g., increasing headway) in order to best manage difficult situations (R. A. Young, 2015). More recently, J. Y. Lee, Gibson, and Lee (2016)[12] reported that when drivers commit errors in a secondary task while driving, the majority will redirect attention to the roadway prior to and post error correction. This suggests that when secondary tasks are to be performed, drivers typically make sure that it is safe to divert attention from the road to deal with the error before engaging in this act. Additionally, post-task checks are made again to update the driver and ensure his or her on-the-road status. In such cases, if a potentially critical event on the road is not perceived during these pre-task checks, secondary task engagement ensues.

The strategies mentioned in the previous paragraph, viz., self-pacing and task demand balancing, serve to avoid danger and minimize accident risk. Serious problems, however, arise when these strategy systems fail, for example, when a driver is not able to sense potential danger (e.g., due to either physical obstruction of the signal, internal noise, or attentional decrements), cannot accurately estimate his or her ability or distraction level (i.e., danger or risk perception) or chooses not to adapt to or ignores environmental demands. Additionally, the prominence of such self-pacing and demand balancing practices in reality can be called into question based on a study published by Vollrath, Huemer, Teller, Likhacheva, and Fricke (2016). Vollrath et al. (2016) reported an observational study where drivers were observed in specified road areas located in three different cities in Germany: Braunschweig, Hannover, and Berlin, and their secondary task engagement was reported while driving and at stand still. In this study, no sig-

---

[12] Although the majority of test persons adopted the strategy as mentioned, the system characteristic of delay (i.e., with or without) also affected the error correction strategy.

nificant differences between driving and stand still status were reported in the percent of drivers performing secondary tasks (e.g., phoning via hand-held and hands-free phone, smart-phone use, eating, drinking and smoking). Therefore, although the percentage of overall drivers engaging in any type of additional distracting activity is generally low (ranging from 10-15% of all drivers observed for all cities tested), those who did take on an additional task to perform, did not generally differentiate between whether they were currently driving (the more difficult situation, possibly requiring danger avoiding performance strategies) or not.

2.5.2. Driver distraction and its measure

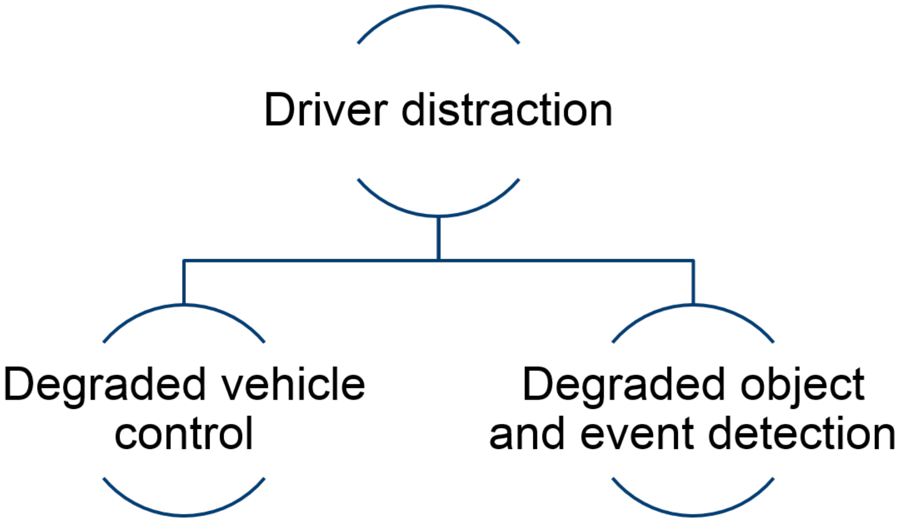


Figure 2.6. – Division of driver distraction manifestations based on Tijerina (2000).

Driver distraction can be generally defined as the inability of a driver of a vehicle to properly attend to driving as a result of some other competing, concrete activity: "... a diversion of attention away from activities critical for safe driving toward a competing activity" (J. D. Lee, Young, and Regan 2009, p. 38; see also Bengler, 2014; Foley, Young, Angell, & Domeyer, 2013; T. A. Ranney, 2008; K. Young, Regan, & Hammer, 2003, 2007). Engström, Victor, and Markkula (2013) defined driver distraction as "... the selection of non-safety-critical schemata that compete with those considered safety-critical, with the consequence that the activation of the safety-critical schemata becomes insufficient" (p. 45). In accordance with these definitions, although the source cause for incidents may differ, they are the result of performing a concrete task in addition to driving with a common outcome: the driving task is not adequately attended to.

Driver distraction can also be defined in terms of its source, effects and manifestations. “While sources of distraction may take many forms, it is helpful to examine distraction in terms of four distinct categories; visual distraction (e.g., looking away from the roadway), auditory distraction (e.g., responding to a ringing cell phone), biomechanical distraction (e.g., manually adjusting the radio volume), and cognitive distraction (e.g., being lost in thought) (T. Ranney, Garrott, & Goodman, 2000, p. 1). Tijerina (2000), however, defined “three broad classes of safety-relevant distraction effects” (Jahn et al., 2005, p. 256): (1) “general withdrawal of attention”, (2) “selective withdrawal of attention” and (3) “biomechanical interference” (Tijerina, 2000, p. 2; Jahn et al., 2005, p. 256). Manifestations of driver distraction can be observed as “degraded vehicle control” or “degraded object and event detection” (both from Tijerina 2000, p. 2; see Figure 2.6). The second class of distraction effect occurs as a result of cognitive or visual load (Jahn et al., 2005) and can manifest as object and event detection detriment. The third class occurs when in-vehicle tasks require physical manipulation or “body shifts” (Tijerina, 2000, p. 2) and can manifest as degraded vehicle control, affecting the “fast and effective execution of maneuvers” (Jahn et al., 2005, p. 256). The first class occurs when “drivers move their eyes away from the road scene” (Jahn et al., 2005, p. 256) and can manifest as both types of distraction manifestation (Jahn et al., 2005; Tijerina, 2000).

In order to access the degree to which certain tasks or devices, etc. can distract the driver on the road, many methods have already been developed such as those already listed in Chapter 1 (occlusion, LCT, eye-tracking and driving metrics). Specifically, with the occlusion method, a surrogate driving scenario is created by permitting intermittent view of a device or interface with which a task is to be performed. The purpose of only allowing intermittent views is to approximate the real life scenario of a driver balancing the demands of driving (i.e., shutter closed, driving task in focus) and that of performing an additional task (i.e., shutter open, device and hand movement may be visually attended to). Comparisons of total task time with and without intermittent viewing as well as the interruptibility of a task may be used to evaluate different systems in terms of their suitability[13] within the vehicle. The ease or difficulty of a task may also be determined through occlusion. Krems, Keinath, Baumann, Gelau, and Bengler (2000), for example, reported based on previous empirical findings that the occlusion method was sensitive enough to decipher between simple and complex tasks despite user or system defined display times. The LCT provides a standardized driving environment where test persons execute a task of interest while driving. The LCT performance itself is based on a test person's ability to detect and respond to given signals (viz., street signs indicating which lane to change to) and maintain lateral control of a virtual vehicle. In eye tracking studies, the degree to which a driver's gaze is diverted from the road is considered a measure of visual distraction. Different in-vehicle devices, for example, can be compared in terms of the degree to which they are associated with glances of a certain length and frequency towards a device of interest and away from the road. Finally, the performance of specific, defined driving metrics, especially those involving lateral and longitudinal control, may also be used to evaluate driver distraction potential. Based on these methods, tasks that are extensively time-consuming and require frequent, long glances away from the road, and/or that are associated with an observable detriment in vehicle control, would be deemed as overly distracting and therefore not suitable to operate in-vehicle while driving.

2.5.3. The problem of quantifying the driver distraction potential of cognitive tasks

Cognitive tasks directly demand human cognition and as a result, attentional performance decrements, such as missed signals or increased response latency, are often observed under such conditions; referred to, as previously mentioned, as a "selective withdrawal of attention" (Tijerina, 2000, p. 2). The techniques discussed in the previous section are only sensitive to distraction potential in so far as defined by the measure's performance metric and lack the sensitivity needed to accurately assess cognitive tasks, associated with different attention-related behavioral decre-

---

[13] As defined in section 3.1.1 of ISO 17287:2003 , suitability is "the degree to which... [use of a system] is appropriate in the context of the driving environment based on compatibility with the primary driving task" (section 3.1.1).

ments. A driver performing a purely cognitive task might appear undistracted according to any of the aforementioned characteristics despite actually being cognitively elsewhere or not available. For example, drivers performing cognitive tasks tend to have lower lane deviations and actually maintain their lane position better than those performing visual-manual tasks (Cooper, Medeiros-Ward, & Strayer, 2013; Engström, Johansson, & Östlund, 2005; Östlund et al., 2004) and tend to maintain their visual focus on the road area in front of the vehicle while performing additional cognitive tasks (Harbluk, Noy, Trbovich, & Eizenman, 2007; Östlund et al., 2004). In Östlund et al. (2004), several experiments were reported where driving performance (longitudinal and lateral vehicle control)[14] was measured while performing either a primarily visual or cognitive secondary task. The visual task used in the reported experiments required participants to visually search for a target, an upward facing arrow, among distractor items (for a detailed description, see Östlund et al., 2004, pp. 12-16). The cognitive task required participants to respond to auditory sounds, specifically, to detect and respond to target sounds (for a detailed description, see Östlund et al., 2004, pp. 16-20). Östlund et al. (2004) reported that the performance of a primarily visual secondary task was associated with a decrease in driving performance (i.e., performance of the primary task), especially in the lateral control of a vehicle, relative to baseline performance and during the performance of primarily cognitive secondary tasks. In other words, cognitive secondary tasks were not associated with the same type of performance decrement as visual manual tasks. Similar findings were reported by Angell et al. (2006) where performing auditory-vocal tasks while driving was associated with a 7% increase of visual focus to the road compared to a 40% decrease during additional visual-manual task performance (p. xxxvi). Correspondingly, Mehler, Reimer, and Coughlin (2012) reported that neither driving performance (viz., *M* and *SD* of forward velocity, and steering wheel reversal rate) nor visual behavior reliably differed between different levels of cognitive task load while driving, despite other metrics (viz., heart rate) being sensitive to such load.

In their SPIDER framework, Strayer and Fisher (2016) review previous findings and conclude that drivers under cognitive demand are particularly poor at evaluating the driving environ-

---

[14] Self-reported driving performance was also measured, however, not relevant in this context.

ment and identifying objects within it, predicting future events and making decisions as well as executing responses based on a given situation. Amado and Ulupinar (2005) as well as Conti, Dlugosch, Schwarz, and Bengler (2013) found that conversing, despite whether the interlocutor was proximate or remote, negatively affected the performance of a concurrent detection task. Strayer and Johnston (2001) compared the reactions of participants on a go/no-go task while performing a pursuit tracking task and listening to the radio or conversing via a hands-free or hand-held cell phone. They found that the probability of missing a stop signal as well as the mean RTs to these signals were significantly higher for participants while conversing than while performing the tracking task alone. Ünal, Platteel, Steg, and Epstude (2013) found that participants were less able to recall experienced events as a result of high demand. Specifically, participants who listened to a radio broadcast while watching a driving video recalled more of the broadcast than participants who listened to the radio and concurrently performed a simulated driving task. This effect was also found for those who drove in simple traffic scenarios relative to more complex ones. Blalock et al. (2014) reported that participants performing a difficult counting task were less accurate in recalling details about their surroundings, especially moving objects, relative to a no-load condition with no additional counting task. Similar to this, Strayer, Cooper, and Drews (2004) also reported that participants were able to recall less objects in a driving scene post-experiment when they were driving and conversing than when just driving.

The distraction potential of cognitive tasks presents a challenge to the already established measurement methods as it cannot be simply observed. Moreover, applying inappropriate measurement methods risk inadequately portraying the driver's mental state and his or her ability or inability to deal with a possibly critical event—having potentially fatal consequences on the road. Therefore, in order to be able to predict performance breakdowns due to concurrent performance of cognitive tasks, to quantify the relevant driver distraction potential, and to avoid it, a valid and accessible measurement method is needed. This thesis focuses on the DRT: a leading option to quantify and estimate the demand cognitive tasks place on the driver and to measure the attentive effects of these tasks. The way the DRT is related to driver distraction can be understood through Figure 2.7: by being able to measure the demands placed on the driver by

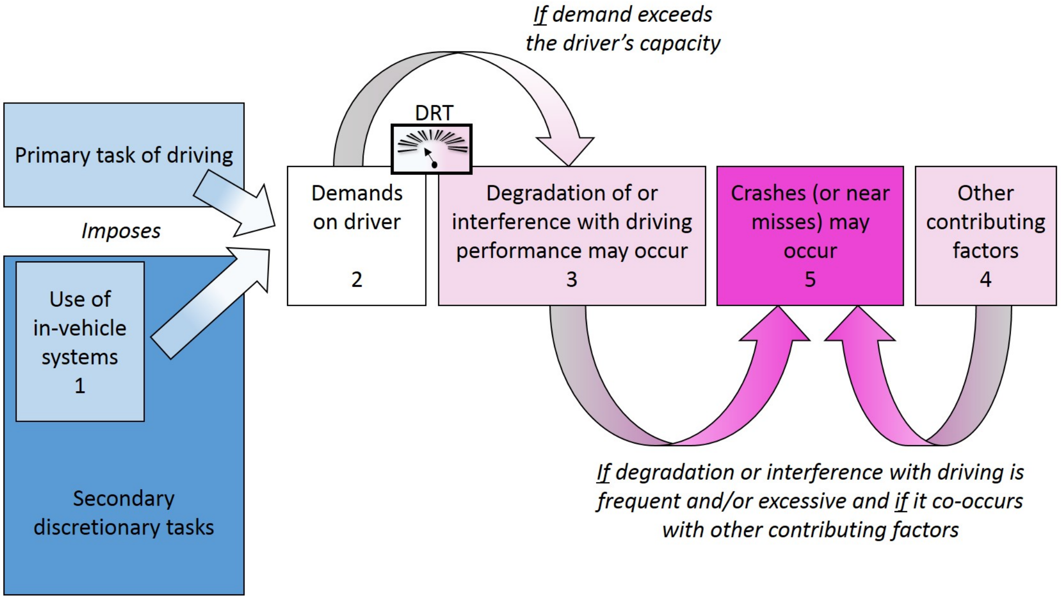


Figure 2.7. – The relation between the DRT measure, driver distraction, and accidents. Figure originally entitled: "Safety Relevance of Driver Workload Metrics" (adapted from Angell et al. 2006, p. xxviii; DRT measure added and format has been changed).

cognitive tasks (blocks 1 and 2) through driving task performance interference or degradation (block 3) in terms of object and event detection, the driver distraction potential of this these cognitive tasks may be inferred. The DRT does not provide a direct measure of distraction, though, as it does not indicate whether and to which degree safety critical activities are ignored. How the DRT measure relates to traffic safety in terms of crash risk (block 5) is discussed in section 2.6.4 on page 44 and in Chapter 5.

2.6. Detection Response Task

The DRT[15] is sensitive to how performing secondary tasks of interest, especially cognitively loading tasks, affect selective attention. As presented earlier in the general overview of the DRT, the performance metrics of the DRT are evaluated as reaction speed (viz., RT) and accuracy (viz., HR or MR). The interpretations of these performance metrics adhere to the following logic: quick RTs indicate that the other concurrently performed task(s) is (are) less demanding than tasks associated with slower RTs; easier tasks do not prevent the participant from responding

---

[15] Note: Although DRT implementation, software and hardware may vary across experiment and laboratory, the experiments discussed here use the same general principle of the DRT and are, therefore, generally compared.

to the DRT signal, whereas more challenging tasks have a higher propensity to do so more frequently. Hence, the speed and rate at which the DRT signals are responded to provides a gauge of task related demand (as per the subsidiary task paradigm discussed in section 2.1.3). Previous studies have shown that DRT performance has decreased as a function of increasing cognitive load, as manipulated by working memory tasks (Conti, Dlugosch, & Bengler, 2014; Conti et al., 2013; Conti, Dlugosch, Vilimek, Keinath, & Bengler, 2012; Engström, Larsson, & Larsson, 2013; Harbluk, Burns, Hernandez, Tam, & Glazduri, 2013; T. A. Ranney et al., 2014; Siam, Isa, Sukardi, Borhan, & Voon, 2014; R. A. Young, Hsieh, & Seaman, 2013), auditory-vocal speech tasks (Conti et al., 2013; Engström, Åberg, Johansson, & Hammarbäck, 2005; Harbluk et al., 2013), a combination of the two aforementioned task types (Harbluk et al., 2013), and tasks requiring mental arithmetic (Bengler, Kohlmann, & Lange, 2012; Diels, 2011; Engström, Åberg, et al., 2005). Consequently, this measurement method has recently been standardized (under ISO 17488:2016) as an applied measure of the "attentional effects of cognitive load" (ISO 17488:2016, title page).

Over time, variants of the DRT have been developed (RDRT, HDRT, and TDRT[16]) and implemented, differing principally in signal modality and presentation location. Different variants were created to account for different perceptual sensitivities and permit flexibility in terms of experimental set-up constraints. The RDRT and HDRT are visual DRTs and deliver their stimuli via LEDs; whereas the RDRT is to be placed in a stationary location, the HDRT moves with the test person's head to control for missed signals or long RTs due to such movements. The TDRT is accomplished by placing a small vibration motor on the shoulder of the participant and delivers its signal by vibrating. Methodological issues aside, which may render one DRT modality more suitable than another, no major measurement differences between the modalities have been reported (Conti et al., 2012; Engström, Johansson, & Östlund, 2005; Merat & Jamson, 2008; T. A. Ranney et al., 2014; Siam et al., 2014). However, since the TDRT has no visual component, it can be considered a more specific, and perhaps exclusive, reflection of a task's demand for cognitive resources (Engström, 2010; Engström, Larsson, & Larsson, 2013;

---

[16] An auditory version of the DRT is also possible, where an auditory signal is to be responded to. An investigation, however, of this auditory DRT was beyond the scope of this thesis.

Hsieh, Seaman, & Young, 2015; R. A. Young et al., 2013).

The history of the DRT method begins with van Winsum, Martens, and Herland (1999), credited with developing an early version of the DRT based on Miura's (1986) investigation of visual attention. In the field experiment reported by Miura (1986), participants drove under conditions of varying difficulty and responded verbally to a detection task (DT[17]). It was found that DT RTs were sensitive to and increased with driving condition demand. Van Winsum et al. (1999) then implemented a DT in a driving simulator as a red square that appeared within the simulated scene and served as the signal. Results from this experiment showed that the DT performance, both RT and MR, increased together with driving task and secondary task load.

Although the DRT is proposed for secondary task assessments, the DRT has also been reported as sensitive to differences in primary task (e.g., driving) load. Bruyas and Dumont (2013) and Jahn et al. (2005) found differing road complexities affected concurrent signal detection such that the more difficult conditions were associated with higher RTs. Similarly, higher RTs have been found for conditions using a driving simulator relative to real-world driving (Siam et al., 2014), as well as conditions without driving as a primary task (viz., "static conditions", Engström, Larsson, & Larsson, 2013). Engström, Åberg, et al. (2005) recorded participants' performance on remote and tactile DRTs while driving either on a motorway, rural or city road, and performing various cognitive tasks. The authors reported that both were sensitive to driving and secondary task differences. On the contrary, Olsson and Burns (2000) used an early form of the RDRT where peripheral signals were presented on the windshield while participants drove on a motorway or country road and performed various visual-manual and cognitive tasks. In this experiment, the DT RTs and HR were sensitive to differences in the secondary tasks but not in the primary task, where the two roads were not discriminated.

2.6.1. The measurement of the DRT

As previously mentioned, the official understanding of the DRT is a tool used to measure the "attentional effects of cognitive load" (see ISO 17488:2016, title page). In the ISO 17488:2016

---

[17] Similar if not identical to the DRT.

document, cognitive load and demand are akin, defined as the "demand for cognitive control imposed by a task" (ISO 17488:2016, section 3.6), drawing on the theory of attention as a control mechanism (discussed in latter two segments of section 2.2) used to select and process certain stimuli. Therefore, the DRT is proposed as a measure sensitive to how human selective attention is affected by a task's demand. An interpretation of the DRT measure can vary according to which model of attention is summoned (as per those models reviewed in section 2.2). An understanding of the DRT based on Wickens' model (section 2.2.1) would contend that as the DRT measures the effects of cognitive demand on attention selection, the DRT reflects changes that occur to the information filtering process and is affected by resource time sharing, induced by task load. An interpretation according to the GAT model (section 2.2.2), might be that when multiple cognitive tasks requiring control are to be performed, attentional control intervenes to serially process these tasks and this is reflected in the DRT metric. As automatic processes do not require attentional control, the DRT metric for these types of tasks might reflect basal processing efficiency. Finally, according to the model proposed by Engström, Victor, and Markkula (2013) and as previously mentioned in section 2.2.3, three different possibilities are considered which would cause multiple task performance interference: a physical reason for miss or response delay, cross talk of activated schemata, or a shared demand for control (Engström, Victor, & Markkula, 2013, pp. 44-45); also valid for the DRT metric.

Previous literature has compared the DRT (or a similar task) to other accessible measures of task load, such as the ERP component P300 and heart rate. The P300 component is a well established marker of event-related cortical activity whose amplitude represents processing capacity (as reviewed in Kok, 2001) and cognitive workload (as reviewed in Hancock, Meshkati, & Robertson, 1985; Wickens, Isreal, & Donchin, 1977). Specifically, Wickens et al. (1977) reported that the P300 component reduced in amplitude as a result of increased task load. Task effects on the component's latency and amplitude are interpreted as the processing time and intensity, respectively (see Kok, 2001, pp. 557-558). Strayer, Turrill, Coleman, Ortiz, and Cooper (2014) investigated a DT similar to the HDRT using the P300 component. Three DT experiments were implemented where cognitive, voice-based tasks were investigated[18]. The DT used in these experiments was a go/no go variant of the HDRT (referred to here as $\mathrm{HDT}_{\mathrm{gng}}$), where participants were to respond to a green light and to ignore a red one. In the first two experiments reported, each task was performed with the $\mathrm{HDT}_{\mathrm{gng}}$ and with (Experiment 1) and without (Experiment 2) a simulated driving task. Additionally, EEG and heart rate measures were also recorded during these experiments. In their first experiment, Strayer and colleagues report a significant main effect of task condition on $\mathrm{HDT}_{\mathrm{gng}}$ RTs and subjective NASA-TLX workload scores. In terms of P300, a significant effect of task load was found for amplitude, but not latency (similar to a study published by Strayer and Drews in 2007, where conversing on a cell phone lead to a reduction of the P300 component, interpreted as a decrease in the encoding process of events possibly relevant to driving). Similar findings were reported in their second experiment, except no significant effect of condition was found on P300 amplitude. Therefore, it seems that with an increase in load (viz., a simulated driving task) and "biological noise", as noted by Strayer et al. (2014, p. 15), the distinction of conditions according to the P300 amplitude disappears. Interestingly, Strayer and colleagues (2014) report no significant effect of task condition on heart rate measures in either experiment–which suggests that the P300 and heart rate, although both related to task load, are at least to some extent sensitive to different aspects of it. Analysis of the driving task, where participants were to follow a vehicle that would brake at unspecified intervals, revealed that both breaking RT and following distance changed as a function of condition. In the third experiment, a field experiment reported by Strayer et al. (2014), the same cognitive tasks were performed in a real vehicle on the road and no additional EEG measures were recorded. Alike the first two experiments, similar effects were reported for the $\mathrm{HDT}_{\mathrm{gng}}$ performance, subjective measures, and heart rate.

Heart rate has been found to increase as a function of cognitive task load. Mehler, Reimer, and Coughlin (2012) recorded and analyzed heart rate (beats per minute [bpm]) differences while participants drove a real vehicle on the road and performed a working memory task (the

---

[18] Specifically: DT baseline, car command, natural listen, synthetic listen, natural listen + compose, synthetic listen + compose, menu high reliability, menu low/moderate reliability, and siri-based interactions (Strayer et al., 2014, pp. 9-11).

audio n-back, discussed in more detail on page 55) in varying levels of difficulty. They found that for the most demanding task (i.e., 2-back), the highest bpm was measured, followed by the mid-difficult (viz., 1-back) and, eventually, easiest (viz., 0-back) task. Heart rate and heart rate variability were also recorded in an experiment reported by Jahn et al. (2005), where the suitability of using these measures to evaluate cognitive load, in addition to an early form of the DRT, was evaluated. In their experiment, taxi drivers drove on more and less difficult roads and performed tasks with an in-vehicle route guidance system presented on a large or small (between-group factor) display. They found that although the DT was sensitive to differences in load, heart rate and heart rate variation (expected to decrease as load increases; see review by Hancock et al., 1985, p. 1112) were not as *selectively* sensitive to just load. The authors concluded that these physiological measures were likely additionally sensitive to physiological responses such as emotional strain and/or physical load. These findings are similar to those reported by Strayer et al. (2014), discussed earlier, where heart rate was also found to not be sensitive to task load in the same way as the DRT.

2.6.2. The DRT reflects objective and subjective characteristics

The DRT reflects both objective, task-based characteristics (i.e., task load), as well as subjective[19] characteristics. This is similar to the concept of *experienced load* expressed by Rouse, Edwards, and Hammer (1993) and reported in de Waard (1996), where the DRT metrics would be sensitive to person-based characteristics (see Mehler, Reimer, & Zec, 2012, for a review). Although this could become problematic when comparing data across laboratories with different test persons, the relevance of inter-individual differences could diminish by focusing on the order of the condition means and how each mean value relates to that observed in other conditions, rather than the absolute mean value itself (ISO 17488:2016, Annex E). However, which inter-individual differences exist as well as the degree to which such idiosyncrasies alter DRT performance metrics, remains to be explicitly empirically evaluated. Strayer, Cooper, Turrill, Coleman, and Hopman (2015), for example, reported age as having a significant effect on mean DRT RT, such

---

[19] As in *influenced by the individual* and not to be confused with self-reported data.

that older groups had longer RTs than younger. Another example was presented by Patten et al. (2006), where it was reported that experienced drivers reacted to a remote DT more quickly than inexperienced drivers while driving on the road. Additionally, from a slightly different perspective, how different characteristics of the DRT, such as signal intensity and placement, affect DRT performance also requires further investigation. An effect of this type was suggested in the experiment presented by Conti et al., 2015, where different TDRT vibration strengths, manipulated through signal frequency, lead to different performance metrics: generally, frequency was inversely related to RT.

2.6.3. The secondary task technique

The DRT utilizes the secondary task technique, specifically, the subsidiary task paradigm, where a test person's residual ability to respond is reflected in its metric, which has distinct advantages and disadvantages. As discussed earlier, according to this technique, a *primary task* is performed together with a *secondary task* and the performance of the latter task provides a value, which quantifies the former (Gopher & Donchin, 1986; O'Donell & Eggemeier, 1986). If similar resources are not used while performing these two tasks, the metric of the secondary task cannot accurately measure the primary task (Gopher & Donchin, 1986; O'Donell & Eggemeier, 1986; Wickens & Hollands, 2000). Ideally, the performance relationship between each task is such that as the primary task increases in difficulty, so too should the performance decrement in the secondary task increase. Ideally and advantageously, the same secondary task can be used to gauge the demand of a variety of primary tasks; despite the diversity of primary tasks, changes in the performance of the secondary task can be easily compared through the same metric (Gopher & Donchin, 1986; O'Donell & Eggemeier, 1986; Wickens & Hollands, 2000). However, problematic assumptions accompanying this technique exist. The following assumptions were identified by Gopher and Donchin (1986) and are discussed here in terms of the DRT.

One assumption is that “subjects are in full control of the allocation of their processing efforts among the two tasks” (Gopher & Donchin, 1986, p. 41-26). Although tasks in a DRT experiment are often referred to as as “primary” and “secondary”, this does not necessarily reflect a participant's ability to prioritize and actual prioritization of each task in a dual or multiple task setting. In a DRT study where multiple tasks were to be performed, instructed prioritization of different tasks within the multiple task set did not produce performance differences, suggesting that participants are not be able to completely maintain a specified task prioritization under loaded conditions (Conti et al., 2014). Another assumption is that "the introduction of the second task does not cause an important change in the nature of the performance conditions of the primary task" (Gopher & Donchin, 1986, p. 41-26). Ideally, task demand is only to be observed in the performance of the secondary task, i.e., the DRT, and despite introducing the DRT, primary task performance should be unaffected, maintained as per baseline. A rather unrealistic assumption, this would imply that by using the DRT, performance differences are seen only in its metrics and the performance on all other concurrent tasks is maintained. When such performance trade-offs occur, the DRT does not a purely reflect demand and may require a more comprehensive interpretation (see O'Donell & Eggemeier, 1986, p. 42-26).

In order to avoid methodological issues with the secondary task technique, O'Donell and Eggemeier (1986) provide a use guideline for "applications of secondary task methodology" (Gopher & Donchin, 1986, p. 41-27). Among the most relevant to the DRT are: (1) to instruct participants to maintain primary task performance as per baseline; (2) to record baseline performances for all tasks to evaluate "intrusion effects" (Gopher & Donchin, 1986, p. 41-27) of the secondary task on the primary task and to evaluate performance changes that occur when shifting from a single task condition to that of a multiple task; and (3) to test different levels of secondary task difficulty to ensure a successful manipulation of demand (Gopher & Donchin, 1986, p. 41-27).

2.6.4. DRT performance and its relation to driver distraction and traffic safety

As mentioned, the DRT does not indicate whether safety critical activities are ignored and therefore does not directly measure distraction. However, it could be argued that the DRT metrics serve to help estimate and predict if and how a signal in the real-world would be reacted to by a driver performing a cognitive task under investigation. Cognitively loading a test person through requiring the performance of a cognitive task, though, is not analogous to distracting him or her. Although they are related concepts, they are distinct from one another: being loaded behind the wheel does not necessitate distraction (Schaap, van der Horst, van Arem, & Brookhuis, 2009). One would suppose that in demanding situations, limited cognitive or control resources are likely to be already engaged to such an extent that the driver could potentially become distracted and no longer able to adequately attend to the driving task and its requirements. Consequentially, the time needed to act in an emergency increases (e.g., late response to traffic event), or even worse, is inhibited (e.g., no response to traffic event). Often though, under difficult conditions, drivers tend to change their behavior and adjust their performance of other tasks (Fuller, 2011; R. A. Young, 2015). Klauer, Dingus, Neale, Sudweeks, and Ramsey (2006) reported that many collisions in the 100-Car field study were due to the driver looking away at a critical moment. In the same vein, if a loaded driver does not react at a certain point, the driver distraction potential and accident risk are surely higher than not; however, accident risk and driver safety are only affected by load given the occurrence of a critical event. Therefore, although the DRT is very sensitive to the demand of cognitive tasks, its metrics exact significance on the road has yet to be determined (this topic is further discussed in Chapter 5). Driving task interference, for example, would be a good candidate for such significance (Shinar, 2015; see also Figure 2.7 on page 37).

In an effort to validate the DRT in terms of traffic safety, Mantzke and Keinath (2015) attempted to show a relationship between DRT and brake RTs in a multiple task condition, but were unsuccessful. In their experiment, participants drove along a simulated road and performed an additional cognitive task, in varying levels of difficulty (4 levels plus baseline). Two driving courses were implemented: one where participants additionally performed the DRT and one where participants had to additionally react to the sudden event of pedestrians on the road. A within-subjects design was implemented to assess whether DRT metrics correlated with braking times to pedestrians. Although the DRT RTs increased along with load, breaking time in reaction to the pedestrian did not. These findings are reminiscent of those reported by Engström

(2011, Paper III), where responding to sudden events (i.e., automatically processed events) were were not affected by cognitive task load. Taking both reports into account, the distinction between a reflection or gauge of task demand, which the DRT provides, and a person's ability to respond in a critical situation, which is not provided by the DRT, is highlighted. Furthermore, the DRT measure is *conditionally* related to traffic safety in terms of accident risk, depending on a driver's capability and or the presence of other factors/events (e.g., a sudden, critical event on the road or other traffic participants compensate/do not compensate for an road event) contributing to an accident (see also Figures on page 37 and on page 30).

Another issue in directly relating DRT results to traffic safety can be found in the experimental implementation of the DRT. In a typical DRT experiment, participants continuously perform a secondary task of interest over a set period of time (1 minute was used in the experiment presented in Chapter 3 and 2 minutes in that presented in Chapter 4) while performing the DRT and optionally, a driving task, simulated or not. Continuous DRT performance ensures that enough measurement points are registered to adequately assess a condition. Arguably, the measurement period could be decreased to a minimum of 5 seconds, corresponding to 1 possible DRT measurement point; this would, however, require a substantial increase in sample size in order to maintain statistical power and for the results to be meaningful. Regardless of the task duration, requiring participants to continuously perform a task on demand is artificial (as per Luce, 1986). As discussed in section 2.5.1, drivers attempt to balance demands according to capability by taking on additional tasks when the situation permits or by adjusting the performance of other concurrent tasks (see Fuller, 2011; Summala, 1996; R. A. Young, 2015; as well as Figure 2.5 on page 30). If, in fact, such strategies are applied during a DRT experiment, the additional effort needed to pace such task performance could also be evident in the DRT performance metrics.

2.7. Summary

In the previous sections, DRT-related concepts were reviewed. The use of signal detection metrics such as RTs and accuracy rates (viz., HR and MR) to measure human performance has a long history dating back hundreds of years. Although initially these metrics were thought of as primarily physical, scientists quickly understood the power of such metrics to gain access to and measure cognitive processes. Recently in psychological research history, signal detection metrics have been implemented in studies investigating cognitive mechanisms such as selective attention and as a way to gauge single or multiple task performance. In dual-task paradigms, RTs can be used according to the loading or subsidiary task paradigm, where RTs are evaluated either for the primary task or the secondary task as a measure of task performance or residual performance capability, respectively. Typically, quicker RTs, lower MRs and higher HRs are indicative of easier task conditions relative to conditions associated with longer RTs, higher MRs and lower HRs.

Attention was described through three prominent and accepted models, each illustrating their own unique interpretation of human attention. According to Wickens as reviewed in section 2.2.1 on page 14, attention can be understood in two ways: as a filter of incoming information and as the fuel sustaining further processing. Multiple task performance predictions based on this model are made in terms of separate and allocatable resources. Tasks requiring the same resources will not be performed together as well as tasks that engage diverse resources. According to the GAT model as reviewed in section 2.2.2 on page 15, attention is conceived as a control mechanism, intervening and aiding performance as needed. This control mechanism is deployed, for example, to perform complex tasks requiring control and to pace processing under circumstances where multiple tasks activate and require the same neural pathways. Finally, the two-dimensional model of attention selection in driving was reviewed in 2.2.3 on page 17. In this model attention is described according to four modes arising from two dimensions. Attention selection occurs based on top-down or bottom-up mechanisms and processes ensue either in a controlled and automatic fashion. Tasks and situations are dealt with through the activation of relevant schemata and those most highly activated, are selected. Here, task performance interference may occur due to a physical impediment or due to an overlap of activated schemata/demand for control.

Sensational and perceptual factors in signal detection were also presented and a general overview of visual and tactile processes was provided. Human peripheral vision is sensitive and visual signals presented in the periphery are, therefore, considered salient and able to direct attention overtly as well as covertly. Tactile signals are often responded to more quickly than visual signals. This is likely due to the fact that tactile signals must be proximal to a person in order for the signal to be sensed. Additionally, tactile signals are also salient and can direct attention and speed responses to events.

Cognitive workload as a complex and multifaceted concept was reviewed and explored through its measure. Research directed at investigating cognitive workload often aims to understand diverse aspects of human cognition and to establish a foundation upon which performance breakdowns may be predicted. Cognitive workload is understood as a state affected and determined by characteristics of an operator, the task under investigation, as well as the performance level required in a given situation. Many methods may and should be used to measure and understand cognitive workload such as primary, secondary, subjective and physiological techniques. Focus was placed on electrophysiological measures due to their relevance in Chapter 4 and frequency analyses of alpha and theta were supported as a viable way of determining the brain state of participants. Previous literature has shown that low alpha amplitudes are associated with high levels of attentive processing and high theta amplitudes are associated with high levels of working memory load and control.

Driving can be conceptualized as the balancing of multiple tasks at once. The driving task was reviewed as the composite of three main subcomponents: primary, secondary and tertiary tasks. Additionally, driving was also conceived of as a control task. In line with Fuller's task capability interface as reviewed in section 2.5.1, a balance between capability and task demand ensures driving task control. This balance is often observed in performance strategies where drivers self-pace task performance. It is, however, uncertain to which degree such self-pacing and task demand balancing strategies are executed in reality. Control is lost when demand outweighs capability. This loss of control is likely to end in a collision, unless a critical event does not occur or if others compensate for the loss of control.

Driver distraction was defined and understood as a decrease in attention towards safety related activities due to another concrete activity. As such, driver distraction can manifest in how a vehicle is controlled or signals are detected. The former type of distraction manifestation, typically used to test tasks or devices requiring visual or physical interaction, can already be readily measured through established and standardized methods. However, tasks that require cognitive interaction require a valid method to access the type of distraction manifestation they are associated with: reduced object and event detection.

Considering the previously presented topics, the DRT was introduced as a method to be used to assess the attentive effects associated with performing cognitive tasks. The DRT is performed in addition to a task or tasks of interest and provides signal detection performance metrics (viz., RTs, HRs, MRs), which can be used to compare different conditions. Previous literature has found that a the ERP component P300, whose amplitude reduces as task load increases, is related to performance of the DRT and both are significantly effected by task load in laboratory settings. In addition to reflecting task-based cognitive load, the DRT also seems to be able to reflect person and stimulus based differences. In terms of relating the DRT measure to driver distraction, it is proposed that the DRT does not directly measure driver distraction, but rather the *potential* a task of interest has to interfere with driving or cognitively distract a driver.

With the theoretical foundation established in this chapter, the groundwork has been laid for the experiments reported in the following two chapters.

Chapter 3

Experiment I on the sensitivity of the DRT

3.1. Overview

The purpose of this experiment was to investigate and compare the sensitivity of different DRT variants to different levels of task load. To this end, three different DRT variants were tested within a single experiment: HDRT, RDRT, and TDRT, together with additional tasks (referred to here as "secondary tasks" [STs]) of varying levels (e.g., easy and hard difficulty levels) and types (e.g., visual-manual, cognitive) of task load. Additionally, whether and how DRT metrics change as a result of being performed concurrently with a simulated driving task were also investigated through the manipulation of dynamic, involving a simulated driving task, and static scenarios, not involving an additional simulated driving task. This experiment was performed in collaboration with industry and reported in Conti et al. (2012). Reflecting changes in the development of the DRT standard, the data have been reanalyzed to fit currently accepted ISO standard definitions. At the time of original publication, the definition of DRT inter-stimulus interval (ISI) was not explicitly defined. Currently, it is understood as the signal onset to onset interval (OnOn) and should be set randomly between 3-5 s. In this experiment, a signal offset to onset interval (OffOn) of 3-5 s was used (corresponding to a OnOn of 4-6 s). The time frame for hits has been updated and redefined as 100 – 2500 ms post-signal onset (rather than 200 –

2000 ms). Different than the current ISO standard, the RDRT used in this experiment was an early version with multiple (4, rather than 1; Figure 1.1 on page 3) LEDs and the HDRT was presented at a distance of ca. 18 cm (see Methods for reference point) rather than 12 - 13 cm from the eye. Additionally, HR[1] was defined here as:

$$HR = \frac{\# Hits}{\# Signals} \tag{3.1}$$

Extending the original publication, supplementary details about the experiment as well as updated and additional analyses are provided in this thesis.

The main research questions addressed in this experiment are:

1. Are all DRTs variants sensitive to differences in task load difficulty?
2. Does this dynamic change when performing a simulated driving task?
3. Does dual-task performance, relative to single task performance, co-vary equally for all DRTs?
4. How many participants are required for each DRT?

3.1.1. Hypotheses

The following hypotheses were made pertaining to the research questions identified on this page:

1. Literature has already established the DRT as being sensitive to different cognitive loads (see section 2.6 on page 37). However, on a perceptual level, tactile detection tasks have been found to be processed more automatically and faster than visual detection tasks (ex. Ng & Chan, 2012). In accordance with the idea that the DRT reflects attention related mechanisms, previous literature found little difference between the performances of different DRT variants (ex. Conti et al., 2012; Engström, Johansson, & Östlund, 2005; Merat

---

[1] The ISO document specifies the HR as defined in Chapter 4, considering premature responses (occurring between 0 - 100 ms post-signal onset). As the DRT device technology at that time ignored these early responses, this HR formula was used.

& Jamson, 2008; T. A. Ranney et al., 2014; Siam et al., 2014). It was hypothesized that although DRT performance differences might exist between DRT variants, all DRT variants would yield longer RTs and lower HRs for hard STs relative to performance during easier STs. Additionally, since the DRT has been reported as being sensitive to cognitive task load (ex. Bengler et al., 2012; Diels, 2011; Engström, Åberg, et al., 2005; Engström, Larsson, & Larsson, 2013; Harbluk et al., 2013; T. A. Ranney et al., 2014; Siam et al., 2014; R. A. Young et al., 2013), the DRT performance was expected to be specifically sensitive to load level differences of cognitive tasks (viz., the n-back, control command task [CC], sentences, and counting tasks [described in Methods starting on this page]), which require no visual and or manual interaction.

2. In line with the logic specified in the previous hypothesis, static conditions, with less tasks to simultaneously perform, were expected to yield quicker RTs and higher HRs relative to dynamic scenarios.

3. Based on Wickens (2002) and Engström, Victor, and Markkula (2013), the TDRT was expected to physically time-share best with the simulated driving task. As both the RDRT and HDRT have a visual component, also required in driving, they were expected to time-share less well. As the RDRT is even more demanding than the HDRT, requiring one to attend multiple possible visual signal locations, the RDRT was expected to time-share least well.

4. In line with previous hypotheses, the TDRT was expected to require the least amount of participants, followed by the HDRT, and finally the RDRT.

3.2. Methods

Methods were already presented in Conti et al. (2012, Methods section). In the following section, however, important methodological details are re-presented and extended in order to foster an understanding of the experiment.

Figure 3.1. – Setup for DRT experiments.

3.2.1. Design

A repeated measures design was implemented to evaluate the sensitivity of the HDRT, RDRT, and TDRT to variations in task load. Task load was varied through two difficulty levels per 5 different STs.

3.2.2. Participants

All participants (N = 18; 9 female, 9 male; all right handed with normal or corrected-to-normal vision) were licensed drivers in possession of a valid license and aged between 19 and 27 years old ($M = 22.33$, $SD = 2.50$).

3.2.3. Apparatus

A small, fixed simulator at the Technical University of Munich (TUM) (Figure 3.1) was used for this experiment. The driver´s seat was centrally located behind an active steering wheel (reconfigurable active yolk from Wittenstein[2]). A 55-inch LCD monitor displayed the simulated highway (SILAB; Würzburg Institute for Traffic Sciences GmbH – WIVW; Veitshöchheim, Germany[3]) in front of the driver and sampling frequency was set at 60 Hz. A separate screen and

---

[2] `www.wittenstein.de/download/control-loading-for-simulation-en.pdf`

[3] `https://wivw.de/en/`

numeric keypad to the right of the driver were used to display and manipulate the surrogate reference task (SuRT). Pre-recorded speech segments used for the control command and sentences tasks were produced with Text Speaker 3.19 (DeskShare Inc., 2000) with the voice character “Hans” (IVONA Software). The German version of the n-back audio was used[4]. The DRT used in this experiment was a USB device developed at the Chair of Ergonomics (TUM). The HDRT consisted of 1 LED mounted to a baseball-type cap, viewed at 18cm (measured from cap-brim intersection to LED). The RDRT was viewed at 80 cm from the participant’s viewpoint and included 4 red LEDs arranged horizontally and spread symmetrically 17° and 32° from center. A small vibrating electric motor was placed on the left shoulder of the participant to implement the TDRT. All DRT responses were registered through a micro-switch Velcroed to the left index finger.

Detection Response Tasks

Three DRTs were used in this experiment: the HDRT, RDRT, and TDRT (Figure 1.1 on page 3). DRT signals remained on for 1000 ms or until button press and participants were instructed to respond as quickly and accurately as possible. The OffOn inter-stimulus interval randomly varied from 3-5 ms.

Simulated driving task

The simulated driving task required participants to travel on a bidirectional highway with two lanes in each direction. No additional traffic was part of the scene and the vehicle had an automatic transmission. Participants were instructed to drive safely at all times and to maintain their position in the right-hand lane, adhering to the speed limit of 80 km/h.

Secondary tasks: n-back, surrogate reference, control command, sentences, counting

These tasks were implemented to induce different levels (viz., easy or difficult) and types (viz., purely cognitive tasks requiring no visual and or manual manipulation or tasks also requiring

---

[4] Available here: `http://agelab.mit.edu/study-tools`

visual-manual interaction) of task load. As such, each ST was performed in two difficulty levels: easy and difficult. These difficulty levels were derived based on an task analysis of the cognitive processes required in each task, as per Jonides et al. (1997). In this expert analysis, tasks were partitioned into cognitive subparts and the best way to manipulate difficulty was decided upon. Five secondary tasks were used in this experiment: an n-back task, an auditory, working memory task requiring verbal input; the surrogate reference task (SuRT), a visual search task requiring manual input; a control command task, a realistic, auditory, working memory task requiring verbal input; a sentences task, a realistic auditory, working memory task requiring verbal input and additionally required sentence comprehension; and a counting task, an auditory, working memory task requiring mental arithmetic and verbal input.

n-back The n-back task used was a system-paced audio task requiring participants to listen to dictated numbers and to remember and repeat these numbers (Mehler, Reimer, Coughlin, & Dusek, 2009). Based on Jonides et al. (1997, p. 471), it was determined that this task requires the following processes: encoding to initiate the processing of information; storage of the information received; rehearsal of information to keep it active in working memory; temporal ordering processes to maintain the correct information sequence; inhibition “to dampen the trace” (Jonides et al., 1997, p. 471) of previously relevant information; and response processes. Difficulty level was manipulated through working memory load; participants had to either repeat the current number (n = 0) or the number stated two steps prior (n = 2). Twenty digits were dictated to participants. Responses were given verbally by the participant and recorded by the experimenter. Participants were instructed to perform this task as accurately as possible and were told to not give up if they lost track of the numbers dictated to them. Performance was evaluated in terms of error percentage calculated based on missed and/or incorrect responses. The DRTs were expected to be sensitive to the difference in difficulty levels for this task.

SuRT The SuRT is based on the Feature-Integration Theory of Attention by A. M. Treisman and Gelade (1980) (see also ISO/TS 14198:2012) and is a user-paced, visual-manual task where participants visually search for a target circle in a display filled with distractors. The target differs from distractors in 1 dimension (size) and task difficulty was manipulated through the ease at which the target was able to be differentiated and selected from distractors. This task is primarily visual-manual and required: visual search to locate the target item; target identification and visual selection to attend to the target; and response processes to manually navigate to the target position and select it. For SuRT easy, the target was a pop-out item, pre-attentive as per A. Treisman (1985), with a large selection field (therefore, less manual interaction needed) and for SuRT hard, no pop-out occurred and the selection field was smaller (therefore, more manual interaction required). Participants were instructed to perform this task as quickly and accurately as possible. The DRTs were not expected to be sensitive to the difference in difficulty levels for this task.

Control command In the CC task, drivers commanded a simulated in-vehicle system to carry out certain functions. The participant's task was to initiate the command sequence according to an instructed goal (e.g., turn on the radio and switch to a certain frequency) and to then follow the system's prompts accordingly. To fulfill the task, participants initiated a command dialog and then were to select an appropriate, goal-relevant option from the verbal menu given out by the system. It was determined that this task required: learning processes to understand command sequencing; encoding to initiate information processing; storage processes to decide on the most goal relevant menu option and to incorporate possible strategies previously used on similar systems; and response processes. Task difficulty was manipulated through communication quality. In the easy condition, the system always understood the participant's commands, while in the difficult level the commands were not always correctly received by the system and the participant had to repeat the given command. The DRTs were expected to be sensitive to the difference in difficulty levels for this task.

Sentences This task was modeled after an experiment reported in Baddeley and Hitch (1974), adapted for the purposes of this experiment. In the sentences task, pre-recorded sentences were played for participants. The task was to actively listen to these sentences, respond to the semantics of the sentence (see Appendix A.1 on page 125 for sentences script), and repeat the sentence using proper syntax. It was determined that this task requires: encoding to initiate information processing; storage of the information received and to incorporate previous experiences or learned information; comprehension processes in order to understand and interpret sentence semantics; and response processes. Task difficulty was manipulated through the need to attend to either one or two aspects of the sentence, and to hold content within working memory for repetition. In the easy condition, sentences read to the participants were syntactically correct, however, may or may not have been semantically correct. In the difficult condition, sentences were not syntactically correct and, again, may or may not have been semantically correct. Participants were tasked with first responding "yes" or "no" to the semantic meaning of the sentence and then to repeat the sentence. When the sentence syntax was incorrect, the sentence was to be repeated correctly. The DRTs were expected to be sensitive to the difference in difficulty levels for this task.

Counting This task was modeled after a task reported by Engström (2011, Paper III) and Bengler et al. (2012). In the easy version, participants were required to count up in steps of two from a three digit number given at the beginning of the counting task. The more difficult version required participants to count backward in steps of seven. This task required: encoding to initiate information processing; storage of the information received and last response; mental computation to arrive at the correct next response; and response processes. Task difficulty was manipulated in the mental arithmetic complexity. Numbers given at the beginning of the counting task served as the starting point from which participants were to count. Within the participant, no two numbers were ever repeated. The DRTs were expected to be sensitive to the difference in difficulty levels for this task.

3.2.4. Procedure

Each participant was introduced to the experiment. The participants were familiarized with the tasks and response methods. Tasks were trained individually and practiced until both the experimenter and participant felt comfortable with the task performance. Recordings were divided into 4 randomized blocks: baseline, HDRT, RDRT, TDRT. Each participant performed 87 conditions: [(5 ST x 2 difficulty levels) x 2 driving conditions x 3 DRTs] + Baselines [(5 secondary tasks x 2 difficulty levels) + 3 DRTs x 2 driving conditions] + Driving baseline. Baselines measurements for all tasks were taken (27 conditions), as well as each DRT with each secondary task with (viz., dynamic scenario; 10 conditions per DRT) and without (viz., static scenario; 10 conditions per DRT) driving. All conditions lasted 1 minute each. For all blocks, static scenario conditions were performed prior to the dynamic scenario conditions. Prior to each dynamic scenario condition, participants were instructed and allowed to adjust their driving speed to 80 km/h and lane position (as per instruction) prior to initiating the next condition.[5] The overall task instruction was to drive safely, when driving was part of the condition, and to perform the tasks as quickly and accurately as possible. The duration of the experiment, including training and breaks, was 2-3 hours.

3.3. Results and discussions

Results are presented separately for each research question addressed. Due to technical errors, the data for .67% (8 conditions) were missing. Little's MCAR was conducted and determined that the missing data were missing completely at random (MCAR), $\chi^2(388) = 37.78,\ p = 1.00$. Since the MCAR mechanism is ignorable and such a low percentage of data was missing, missing data were remedied with single mean value imputation (see A. R. T. Donders, van der Heijden, Stijnen, and Moons 2006). For RT analysis, only the RTs of hits (a first reaction given between 100 – 2500 ms post-signal to a signal) were evaluated. Accuracy was also evaluated: missed signals (signals not responded to within 100 – 2500 ms post-signal onset) were assessed in a hit rate analysis. Outliers and normality were assessed. Unless indicated otherwise, no data were excluded from analysis.

---

[5] Note: After each condition was completed, participants were asked to report subjectively on the scene just completed. They did this by completing the mental workload subscale of the NASA-TLX. The evaluation of this data was not considered relevant to the current assessment and, therefore, not included here.

3.3.1. HDRT, RDRT, and TDRT are sensitive to differences in task load

In order to address the research questions indicated on page 51, six (1 per DRT per scenario) one-way repeated measures ANOVAs were conducted to determine whether there were statistically significant differences in DRT RTs for each condition's task load. The HRs were split similarly and evaluated with a non-parametric alternative: the Friedman test with Bonferroni corrected pairwise comparisons (SPSS 22). Significant pairwise comparisons of interest, i.e., that between easy and difficult levels of a ST, are considered and elaborated with mean differences and 95% CI. Pairwise comparisons between STs and baseline measures (DRT for static scenarios, DRT with simulated driving for dynamic scenarios), are also reported.

HDRT

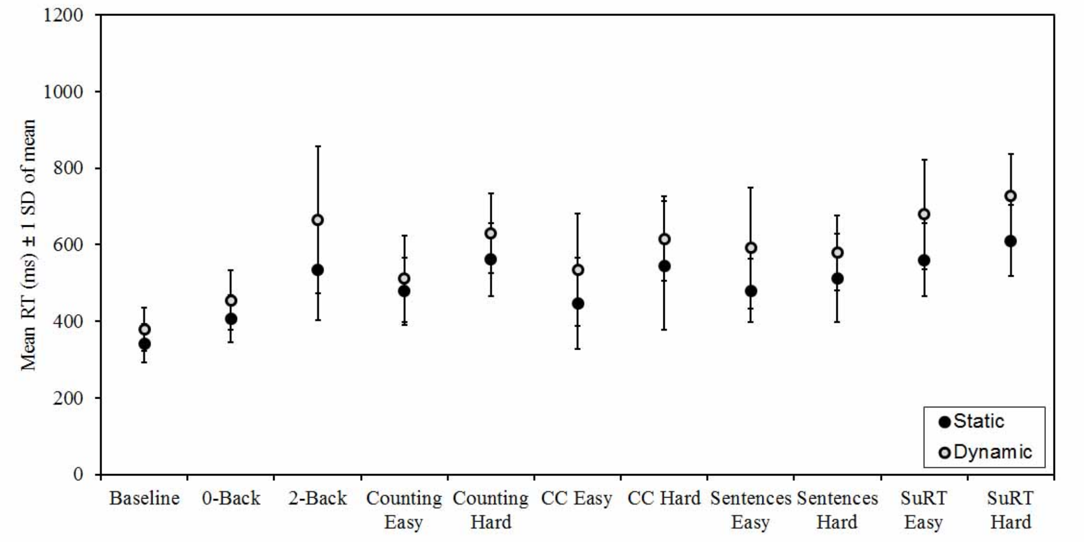


Figure 3.2. – Mean HDRT RTs (ms) ± 1 *SD* per condition (static = without additional simulated driving task; dynamic = with additional simulated driving task).

The *M*s and *SD*s of HDRT RTs are plotted in Figure 3.2 and these values as well as those of the HR (*Mdn*, minimum and maximum values[6]), can be found in Table 3.1 on the following page.

---

[6] Minimum and maximum values are provided as supplementary to *Mdn* values.

	Mean RTs (ms)		Median HR (%)		Minimum HR (%)		Maximum HR (%)	
Condition	Static	Dynamic	S	D	S	D	S	D
Baseline	342.66 (49.77)	379.19 (56.26)	100	100	100	100	100	100
0-Back	406.57 (61.87)	455.46 (77.39)	100	100	100	92	100	100
2-Back	534.90 (131.52)	665.89 (192.12)	100	100	58	33	100	100
Counting (+2)	478.99 (87.38)	511.15 (112.66)	100	100	83	82	100	100
Counting (-7)	561.28 (95.28)	629.84 (103.79)	100	100	77	75	100	100
CC easy	446.12 (118.67)	534.84 (147.13)	100	100	100	67	100	100
CC hard	544.90 (168.24)	615.59 (110.58)	100	100	83	71	100	100
Sentences easy	480.33 (83.35)	591.03 (157.93)	100	100	83	75	100	100
Sentences hard	513.14 (114.89)	579.22 (97.61)	100	100	82	75	100	100
SuRT easy	560.83 (95.34)	679.08 (143.86)	100	100	92	75	100	100
SuRT hard	610.60 (92.21)	727.18 (109.59)	100	100	91	83	100	100

Table 3.1. – HDRT values for static (S) and dynamic (D) scenarios. RTs reported as: *M* (*SD*) and HRs reported as: *Mdn*, minimum and maximum.

Static scenario

Reaction time Mauchley's test of sphericity indicated that the assumption of sphericity had not been violated, $\chi^2(54) = 67.96,\ p = .130$. Task load elicited statistically significant changes on mean HDRT RTs, $F(10, 170) = 16.98,\ p < .001$, partial $\eta^2 = .500$. Bonferroni corrected post-hoc analyses were performed to test for significant differences in mean HDRT RTs between each easy and difficult level of ST. Mean RTs significantly increased from the easy to difficult n-back task with a mean difference of 128.34 (95% CI $[26.27$ to $230.40])$ ms, $p = .006$. Mean RTs also increased from the easy to difficult counting tasks with a mean difference of 82.29 (95% CI $[23.10$ to $141.47])$ ms, $p = .002$. All other difficulty level comparisons were not significant. Additionally, the baseline was significantly different than all other conditions, $p \leq .001$, except for the easy n-back and CC tasks.

Hit rate A Friedman test was run to determine whether there were differences in HDRT HR distributions for different task loads. Significant differences were found $\chi^2(10) = 29.79,\ p = .001$. Pairwise comparisons with a Bonferroni correction for multiple comparisons were performed to assess differences between easy and difficult levels; no significant differences were found. Baseline comparisons were also not significant.

Dynamic scenario

Reaction time Mauchley's test of sphericity indicated that the assumption of sphericity had not been violated, $\chi^2(54) = 61.48,\ p = .281$. Task load while performing a simulated driving task elicited statistically significant changes in mean HDRT RTs, $F(10, 170) = 19.19,\ p < .001$, partial $\eta^2 = .530$. Bonferroni corrected post-hoc analyses were performed to test for significant differences in mean HDRT RTs between each easy and difficult levels of ST. Similar to the results of the static scenario, mean RTs significantly increased from the easy to difficult n-back tasks with a mean difference of 210.43 (95% CI $[61.39$ to $359.48])$ ms, $p = .002$. Mean RTs also increased from the easy to the difficult counting tasks with a mean difference of 118.70 (95% CI [15.77 to 221.62]) ms, $p = .013$. All other difficulty level comparisons were not significant. Additionally, the baseline was significantly different than all other conditions, $p \leq .01$, except for the easy n-back task.

Hit rate A Friedman test was run to determine whether there were differences in HDRT HR distributions for different task loads. Significant differences were found $\chi^2(10) = 21.17$, $p = .020$. Pairwise comparisons with a Bonferroni correction for multiple comparisons were performed to assess differences between easy and difficult levels; no significant differences were found. Baseline comparisons were also not significant.

RDRT

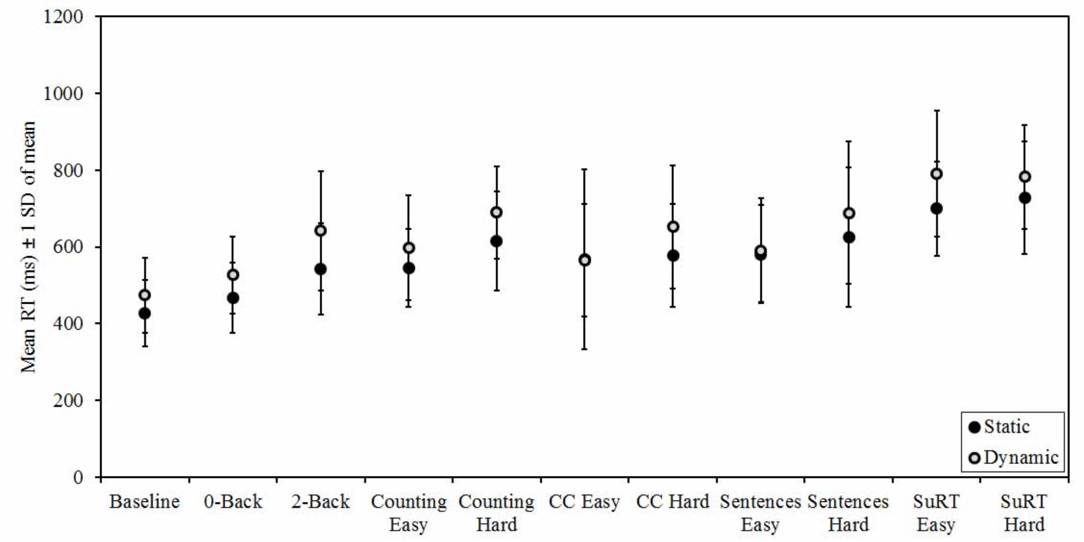


Figure 3.3. – Mean RDRT RTs (ms) ± 1 *SD* per condition (static = without additional simulated driving task; dynamic = with additional simulated driving task).

The *M*s and *SD*s of RDRT RTs are plotted in Figure 3.3 and these values as well as those of the HR (*Mdn*, minimum and maximum values), can be found in Table 3.2 on the next page.

Static scenario

Reaction time Mauchley's test of sphericity indicated that the assumption of sphericity had been violated, $\chi^2(54) = 77.45$, $p = .031$, therefore degrees of freedom were corrected using

Condition	Mean RTs (ms)		Median HR (%)		Minimum HR (%)		Maximum HR (%)	
	Static	Dynamic	S	D	S	D	S	D
Baseline	427.57 (86.43)	474.02 (98.65)	100	100	67	75	100	100
0-Back	467.23 (90.79)	527.46 (100.37)	100	100	92	82	100	100
2-Back	543.24 (119.76)	642.06 (155.91)	100	100	91	67	100	100
Counting (+2)	544.77 (101.93)	598.18 (135.79)	100	100	91	82	100	100
Counting (-7)	616.34 (129.31)	689.89 (120.86)	100	91.67	82	75	100	100
CC easy	568.02 (234.18)	565.63 (147.04)	100	100	75	50	100	100
CC hard	578.84 (133.88)	652.74 (160.19)	100	100	100	60	100	100
Sentences easy	581.29 (127.57)	591.08 (135.81)	100	100	91	82	100	100
Sentences hard	626.39 (182.10)	689.57 (185.03)	100	100	82	83	100	100
SuRT easy	700.46 (123.09)	790.55 (164.37)	100	90.91	82	33	100	100
SuRT hard	727.79 (147.01)	783.20 (135.34)	90.91	81.82	55	42	100	100

Table 3.2. – RDRT values for static (S) and dynamic (D) scenarios. RTs reported as: *M* (*SD*) and HRs reported as: *Mdn*, minimum and maximum.

Greenhouse-Geisser estimates of sphericity ($\varepsilon = .488$). Task load elicited statistically significant changes in mean RDRT RTs, $F(4.88, 82.99) = 10.01$, $p < .001$, partial $\eta^2 = .371$. Bonferroni corrected post-hoc analyses were performed to test for significant differences in mean RDRT RTs between each easy and difficult level of ST, however, no difficulty level comparisons were significant. All comparisons with the baseline, however, were significant, $p < .05$, except those with the easy n-back and CC tasks.

Hit rate A Friedman test was run to determine whether there were differences in RDRT HR distributions across different task loads. Significant differences were found $\chi^2(10) = 43.96$, $p < .001$. Bonferroni corrected pairwise comparisons were performed to assess differences between easy and difficult levels; no significant relevant differences were found. Baseline comparisons were also not significant.

Dynamic scenario

Reaction time Mauchley's test of sphericity indicated that the assumption of sphericity had not been violated, $\chi^2(54) = 63.77$, $p = .219$. Task load while performing a simulated driving task elicited statistically significant changes in mean RDRT RTs, $F(10, 170) = 14.07$, $p < .001$, partial $\eta^2 = .453$. Bonferroni corrected post-hoc analyses were performed to test for significant differences in mean RDRT RTs between each easy and difficult level of ST, however, no difficulty level comparisons were significant. All comparisons with the baseline, however, were significant, $p < .05$, except those with the easy n-back, counting, CC, and sentences tasks.

Hit rate A Friedman test was run to determine whether there were differences in RDRT HR distributions across different task loads. Significant differences were found $\chi^2(10) = 62.66$, $p < .001$. Bonferroni corrected pairwise comparisons were performed to assess differences between easy and difficult levels; no significant relevant differences were found. Baseline comparisons were only significant for the hard and easy SuRT, $p = .003$ and $p = .042$, respectively.

TDRT

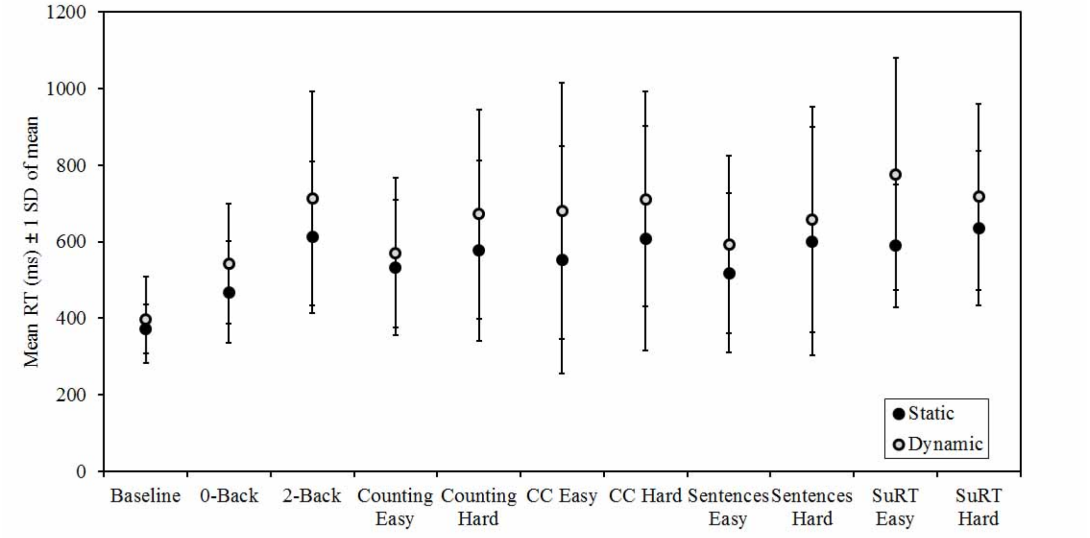


Figure 3.4. – Mean TDRT RTs (ms) ± 1 *SD* per condition (static = without additional simulated driving task; dynamic = with additional simulated driving task).

The *M*s and *SD*s of TDRT RTs are plotted in Figure 3.4 and these values as well as those of the HR (*Mdn*, minimum and maximum values), can be found in Table 3.3 on the following page.

Static scenario

Reaction time Mauchley's test of sphericity indicated that the assumption of sphericity had been violated, $\chi^2(54) = 164.50,\ p < .001$, therefore degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity (ε $= .273$). Task load elicited statistically significant changes in mean TDRT RTs, $F(2.73, 46.37) = 6.047,\ p = .002,$ partial $\eta^2 = .262$. Bonferroni corrected post-hoc analyses were performed to test for significant differences in mean TDRT RTs between each easy and difficult variant of ST. Mean RTs significantly increased from the easy to hard CC task with a mean difference of 56.64 (95% CI $[5.67$ to $107.61]$) ms, $p = .019$. All other difficulty level comparisons were not significant. Additionally, the baseline was significantly different than all other conditions, $p < .05$, except for the easy n-back, and the easy and hard CC and sentences tasks.

Condition	Mean RTs (ms)		Median HR (%)		Minimum HR (%)		Maximum HR (%)	
	Static	Dynamic	S	D	S	D	S	D
Baseline	372.47 (64.74)	396.46 (113.57)	100	100	82	100	100	100
0-Back	468.19 (133.11)	543.28 (155.86)	100	100	92	85	100	100
2-Back	612.91 (198.22)	713.64 (279.59)	100	96.15	67	50	100	100
Counting (+2)	533.74 (176.82)	571.65 (195.02)	100	100	83	83	100	100
Counting (-7)	577.38 (235.72)	672.10 (273.20)	100	100	83	55	100	100
CC easy	552.44 (297.78)	680.28 (334.99)	100	100	60	56	100	100
CC hard	609.08 (293.73)	711.84 (280.45)	100	100	71	63	100	100
Sentences easy	518.37 (208.17)	593.27 (232.25)	100	100	91	73	100	100
Sentences hard	601.48 (298.76)	659.21 (294.50)	100	95.83	91	55	100	100
SuRT easy	589.45 (160.12)	777.06 (303.60)	100	100	92	73	100	100
SuRT hard	636.13 (201.43)	717.53 (243.09)	100	100	92	67	100	100

Table 3.3. – TDRT values for static (S) and dynamic (D) scenarios. RTs reported as: *M* (*SD*) and HRs reported as: *Mdn*, minimum and maximum.

Hit rate A Friedman test was run to determine whether there were differences in TDRT HR distributions for different task loads; no significant differences were found $\chi^2(10) = 15.80,\ p = .105$.

Dynamic scenario

Reaction time Mauchley's test of sphericity indicated that the assumption of sphericity had been violated, $\chi^2(54) = 80.774,\ p = .017$, therefore degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity ($\varepsilon = .518$). Task load while performing a simulated driving task elicited statistically significant changes in mean TDRT RTs, $F(5.18, 88.05) = 9.07,\ p < .001,$ partial $\eta^2 = .348$. Bonferroni corrected post-hoc analyses were performed to test for significant differences in mean TDRT RTs between each easy and difficult level of secondary task, however, no difficulty level comparisons were significant. All comparisons with the baseline, however, were significant, $p < .02$.

Hit rate A Friedman test was run to determine whether there were differences in TDRT HR distributions for different task loads. Significant differences were found $\chi^2(10) = 27.44,\ p = .002$. Bonferroni corrected pairwise comparisons were performed to assess differences between easy and difficult secondary task variants; no significant differences were found. Baseline comparisons were also not significant.

Discussion

This portion of the analysis aimed to determine whether the different DRT variants were sensitive to differences in task load and to answer research questions 1 and 2 identified on page 51. Task load was manipulated by condition: 5 different tasks were performed in two different difficulty levels. Since the DRT can be used in an in-vehicle setting together with an additional driving task, task load was tested both with and without an additional simulated driving task. Six repeated measures ANOVAs were used to assess the effect of task load on mean DRT RTs and six Friedman tests were used to assess the effect on mean DRT HRs. As the difference between task difficulty pairs was of primary interest, where applicable, these are focused on in this discussion. Effect size interpretations reflect the classifications (small = .0099, medium = .0588, and large = .1379) specified by J. Cohen (2009, pp. 284-288; see also Richardson, 2011).

In both the static and dynamic scenarios, the HDRT was sensitive to differences in task load and had large effect sizes. In the RT analysis, the difficulty levels of the n-back and counting tasks were differentiated according to their presumed load: the hard level (i.e., the more demanding task) had longer mean RTs than the easy level (i.e., the less demanding task). The baseline measures were also significantly different than all conditions, save the easy n-back and CC (static only) conditions. For the HR analyses of both the static and dynamic scenarios, the distributions of all conditions were dissimilar, however, no additional information was provided through the statistical post-hoc tests and baseline comparisons.

The RDRT was also sensitive to differences in task load for both static and dynamic scenarios. The effect size for the dynamic scenario was larger than that of the static scenario. Both effect sizes were, however, large. Differences in task difficulty levels were distinguished neither in the RT nor HR analyses. Differences compared to the baseline measurements, however were found. For the RT analysis, the static baseline was significantly faster than all other conditions except the easy n-back and CC tasks. Additionally, the dynamic baseline was significantly faster than all conditions except the easy n-back, counting, CC, and sentences tasks. The HR analysis of both static and dynamic scenarios revealed that HR distributions varied across task load, however, only the HR for the static baseline was significantly higher than for the easy and hard SuRT tasks.

Similar to the other two DRTs, the TDRT was also sensitive to differences in task load for static and dynamic scenarios. The effect size for the dynamic scenario was larger than in the static scenario. Both effect sizes were, however, large. In the static scenario, the mean DRT RTs were not sensitive to differences in task difficulty levels, except for the CC task. For comparisons with the baseline, no significant difference was found between the static baseline and the easy n-back, CC, sentences tasks as well as with the hard sentences and CC tasks. Also, the HR distribution did not differ according to task load. Further analyses of the dynamic scenario showed that no difficulty levels were able to be distinguished in either the RT or HR analyses. Mean baseline RTs for the dynamic scenario were, however, significantly faster than all other conditions.

In line with previous studies (presented in section 2.6: Conti et al. 2014, 2013, 2012; Engström, Larsson, and Larsson 2013; Harbluk et al. 2013; T. A. Ranney et al. 2014; Siam et al. 2014; R. A. Young et al. 2013), all DRTs were found to be sensitive to differences in task load as graphically depicted in Figures 3.2, 3.3 and 3.4. As others have reported (e.g., T. A. Ranney et al., 2014), the DRT RTs were more sensitive than the HRs, which is most likely due to a ceiling effect. Because HR are upward limited, this current analysis supports using RTs as the primary performance metric of the DRT. For this reason, the remainder of the discussion will focus on the RT analyses.

In response to the first research question identified on page 51, of all DRTs, the HDRT had the largest effect sizes and was consistently sensitive in the static and dynamic scenarios as well to the differences in the demand of the n-back and counting task difficulty levels. As these two secondary tasks were assessed as being purely cognitive, this is a good indication that the HDRT is sensitive to tasks requiring cognitive processes, especially those involving working memory and mental computation. Additionally, for all DRTs, the SuRT difficulty pair did not yield significantly different performances, in line with the claim that the DRT is diagnostic of and specific to cognitive demand and not visual and or manual demand of a secondary task (e.g., Hsieh et al., 2015; R. A. Young et al., 2013, for the TDRT).

A challenge to the current experiment, however, is the inhomogeneous sensitivities of the DRT variants to the demand of cognitive STs. Often, easy levels of tasks were discriminated neither from the respective baseline measures nor from their task difficulty partner. For example, the CC task difficulty levels were only discriminated under the TDRT static condition and differences in the difficulty levels of the n-back and counting tasks were only observed for both the static and dynamic HDRT. The difficulty levels for the sentences task were not discriminated by any DRT variant. Although all secondary tasks endured a task analysis, it cannot be sure whether some tasks truly demanded cognitive processes to an extent that was measurable by the DRT. Although this is not likely the case for the n-back and counting task, as DRT differences have been observed for these STs in previous literature (n-back: e.g., Harbluk et al., 2013; counting: e.g., Bengler et al., 2012), the CC and sentences tasks as implemented in this experiment have no such support and would require additional testing.

The RDRT was the least sensitive of the DRT variants, which is likely due to the eccentricity of the task; requiring test persons to attend many LEDs at once, head movement artifacts could have been present in the data, which confound a clear effect specific to task load. The lack of sensitivity of the TDRT, however, is somewhat unusual as the TDRT has been reported as even more sensitive than other DRT variants (e.g., Harbluk et al., 2013). Differences in the vibration motor used in this experiment could explain this divergence in findings (as per Conti et al., 2015). It is also possible that the TDRT and RDRT are simply less sensitive to minute differences in cognitive task load and were generally less effected by the task demand manipulated in this experiment. Alternatively, the statistics including the alpha corrections used could have altered the (in)sensitivity of the TDRT and RDRT. Of course, the lack of distinction between the difficulty levels of some tasks could have also been a random finding and therefore not associated with any real systematic effect of DRT variant.

In response to the second research question identified on page 51, for static and dynamic conditions, the effect sizes for all DRT variants were larger when performed together with the simulated driving task. For the HDRT and RDRT, sensitivities to task difficulty levels were consistent between static and dynamic conditions. However, the sensitivity of the TDRT to the CC task difficulty levels observed under the static condition was lost when performed with an additional driving task. This finding suggests that the TDRT has a dynamic relationship with the driving condition: the formerly found effect is lost when task load generally increases as a result of performing an additional task (viz., simulated driving). With the exception of the TDRT, static and dynamic condition sensitivities were always similar.

In conclusion, all DRTs were sensitive to differences of task load and mostly specific to cognitive load manipulations, though often statistical significance of post-hoc tests was lacking. The strongest and most reliable task difficulty level comparisons were provided by the counting task, n-back and SuRT, the latter of which did not, as expected, produce differences in any of the

DRT RTs. DRT RTs were more sensitive than HRs, supporting the practice of using them as the primary evaluation metric. The HDRT provided the most sensitive and reliable metric. Finally, static and dynamic conditions generally coincided in terms of sensitivity to ST difficulty levels.

3.3.2. POC Analysis: Task performances co-vary differently for each DRT

Condition	RMSE
Baseline Driving	1.46 (0.91)
Driving + HDRT	2.13 (1.76)
Driving + RDRT	3.20 (2.06)
Driving + TDRT	2.22 (1.58)

Table 3.4. – RMSE of driving speed (km/h) per condition, reported as: *M* (*SD*).

The aim of this analysis was to determine whether the DRT variants yielded different performance variances when performed together with a simulated driving task and to answer the third research question identified on page 51. Performance Operating Characteristics (POC) graphically display the performance of (at least) two tasks and how they timeshare (Boff & Lincoln, 1988b; Grondin & Macar, 1992; Navon & Gopher, 1979; Norman & Bobrow, 1975, 1976; O'Donell & Eggemeier, 1986; Tsang & Vidulich, 2006). To perform this analysis, the performance of each task of interest performed alone is plotted on each axis. Concurrent performance of each task together is then plotted, indicating performance changes occurring when multiple tasks are performed together. POC divergences between the DRT variants would be indicative of DRT variant induced strategy differences, reflecting performance limitations. As DRT RTs resulted as the more sensitive DRT metric, these were used in the current POC analysis. Additionally, driving performance was indicated by the root mean square error of speed (RMSE; Table 3.4), calculated using the difference between a participant's speed (*Speed*) and the instructed speed (80 km/h; *ExpectedSpeed*) and the count of values per condition (*n*):

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} (Speed_i - ExpectedSpeed)^2}{n}} \tag{3.2}$$

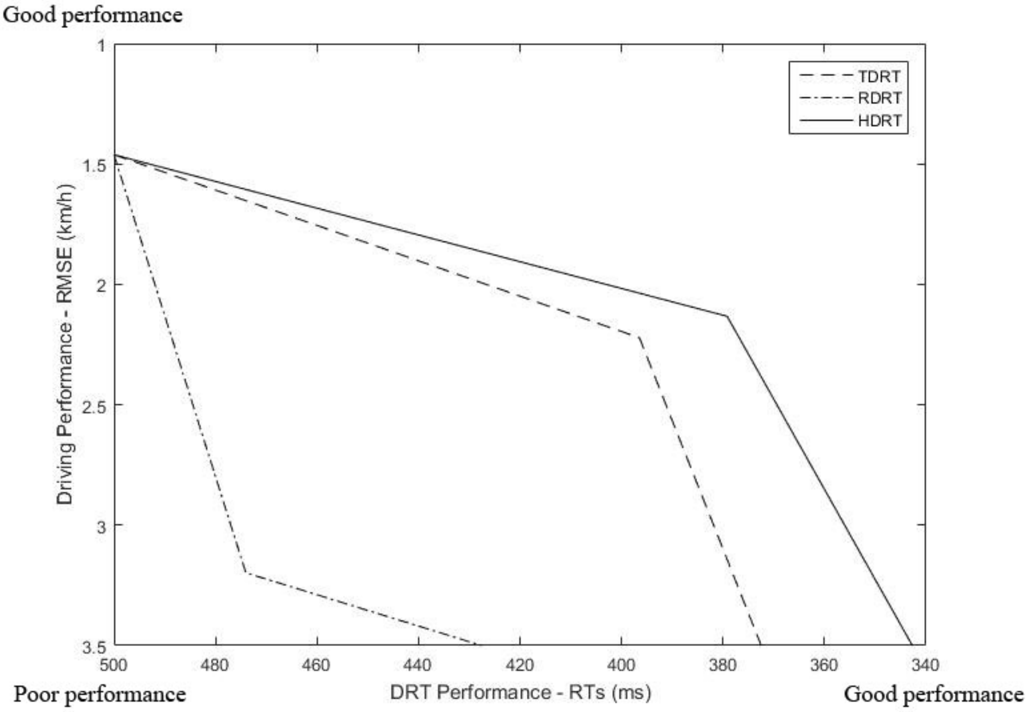


Figure 3.5. – POCs of overall DRT performance (mean DRT RTs [ms]) and driving performance (mean RMSE [km\h]) for TDRT, RDRT, HDRT.

This driving metric was chosen over a lane keeping metric as it has been shown that the effect of cognitive load on lane keeping is minimal (discussed earlier: Cooper et al., 2013; Engström, Johansson, & Östlund, 2005; Östlund et al., 2004). The POC analysis can be found in Figure 3.5.

Discussion

It was hypothesized based on Wickens (2002), that the TDRT would time-share best with the simulated driving task due to a lack of visual overlap present in the RDRT and HDRTs. Additionally, the HDRT was expected to time-share better with driving than the RDRT due to task simplicity. As can be seen in Figure 3.5, performance changes can be observed for all tasks relative to their respective baselines, meaning that performance of all tasks changed as a result of concurrent performance. The HDRT baseline had the fastest mean RTs, followed by the TDRT, and finally the RDRT. Although the means of the HDRT and TDRT varied, these two DRT variants were similar in terms of how their performance changed as a result of simultaneous performance with the simulated driving task. When performed together with the simulated driving task, HDRT, TDRT, and driving performance only slightly worsened. The RDRT, however, yielded the most drastic change when performed together with the simulated driving task. When performed together with the simulated driving task, RDRT and driving performance worsened the most. One reason for this extreme detriment in performance could be due to the nature of the RDRT as implemented in this experiment. The RDRT had a stimulus eccentricity factor not present in the other DRTs. As such, the task is more difficult to perform, requiring a different strategy in order to perform it and attend to the multiple LEDs. This is evident not only in the dual-task performance, but also in the simpler baseline condition. Interestingly, based on the reviewed literature (ex. Ng & Chan, 2012), if primarily sensational/perceptual factors strongly influenced RTs, the TDRT should have produced RTs faster than the visual DRTs. Although this was the case when compared to the RDRT, it was not so for the HDRT. Overall, the results of this analysis suggest that the HDRT and TDRT are less intrusive and cause less performance detriment in other tasks than the RDRT.

3.3.3. Statistical Power Analysis: How many participants are required for each DRT variant?

The following assessment was executed post-hoc to determine the minimum sample size values needed to reliably discriminate between different levels of task load among the HDRT, TDRT, RDRT and to answer the fourth research question identified on page 51. Since the n-back and SuRT were among the tasks that provided the most reliable results in the analysis presented in section 3.3.1, the current analysis was performed on these tasks difficulty pairs as well as baseline measurements.

In null hypothesis significance testing (NHST), an experimenter typically hypothesizes that a difference can be found in a dataset according to some condition or manipulation. This hypothesis is known as the alternative hypothesis: H_A. The counter hypothesis is the null hypothesis, H_0, where no difference between the groups is to be found. There are two major errors that may occur in NHST: Type I (false positives; rejecting the H_0 when it is true and accept the H_A) and Type II (false negatives; fail to reject the H_0 when it is not true, and reject the H_A) errors

(B. H. Cohen, 2013b, p. 143). When a NHST test is performed, a significance criterion (conventionally, .05) determines the fate of the null and alternative hypotheses. This significance criterion is based on α, which is the "long-term" (Dienes, 2008, p. 62) probability of a Type I error (Dienes, 2008). By comparing the *p* value, the probability of the observed statistic if the H_0 were true, to the significance criterion, the H_0 is either rejected ($p < .05$) or failed to be rejected ($p > .05$). This means that although NHST control well for Type I errors, the probability of Type II errors (called β) is not considered as serious a problem. To exemplify this, Liu (2014, p. 19) uses the idea of imprisoning an innocent person (viz., a Type I error), which is argumentatively worse than letting a criminal loose (viz., a Type II error). As one could imagine situations where a false negative would be more serious than a false positive, ideally both errors need to be controlled for. In statistical power analysis, the probability that the H_0 will be rejected when it is false is calculated (B. H. Cohen, 2013b; J. Cohen, 1992, 2009). Specifically, the statistical power of a test is "the probability of not making a Type II error" (B. H. Cohen, 2013b, p. 237), the residual or opposite of β (so, $1 - \beta$), and is considered minimal at .70 (B. H. Cohen, 2013b, p. 242) and sufficient at around .80 (B. H. Cohen, 2013b, p. 242; Field, 2013, p. 69); resulting in a Type II error rate of .30 and .20, respectively. If a test results as having a power equal to or greater than .80 and given an effect does actually exist, it is very likely that it will be found (Field, 2013, p. 69).

In order to conduct a power analysis, J. Cohen (2009) specifies that statistical power depends on the significance criterion, result reliability, and effect size. Sample size needed for an experiment can also be determined through this analysis (Liu, 2014), given α, power, and effect size. Effect size is a measure of the "...magnitude of the difference between groups" (Sullivan & Feinn, 2012, p. 279). Practically, there are a two relatively simple ways to calculate the minimum required sample size needed for a study through already existing software: G*Power (Faul, Erdfelder, Lang, & Buchner, 2007) and the R package 'pwr' (Champely, Ekstrom, Dalgaard, Jill, & De Rosario, 2015).

In this analysis, G*Power was used with an "a priori" type of power analysis. "An a-priori power analysis can provide an indication of the average sample size a study needs to observe

H_0 ($\mu_{diff} = 0$)	$N_{minimum}$ α = .05	α = .01	α = .001	Cohen's d_z
TNN, TDN	80	119	174	.318
TNBE, TNBH	14	20	30	.845
TNVE, TNVH	72	107	156	.337
TDBE, TDBH	14	20	30	.846
TDVE, TDVH	127	188	276	.251
RNN, RDN	20	30	44	.673
RNBE, RNBH	19	28	41	.696
RNVE, RNVH	305	454	665	.161
RDBE, RDBH	12	18	27	.894
RDVE, RDVH	5129	7633	11159	.039
HNN, HDN	10	15	22	1.036
HNBE, HNBH	8	12	18	1.188
HNVE, HNVH	38	56	83	.472
HDBE, HDBH	7	10	15	1.334
HDVE, HDVH	54	80	117	.392

Table 3.5. – Required sample size to reject the H_0 (viz. the mean [μ] difference between the paired values is equal to 0) given desired power (.8) and effect sizes (Cohen's d_z) calculated from paired-samples *t*-tests for different levels of α. Values are based on 1-minute long scenes.

a statistically significant result with a desired likelihood" (Lakens, 2013, p. 1). Although specific hypotheses were made regarding the direction of the differences between task difficulty pairs, two-sided tests were used as they are more conservative than one-sided tests (B. H. Cohen, 2013a, pp. 144-146) and were, therefore, more fitting to the sample size estimation provided by the power analysis. Fifteen paired-samples t-tests were run for each task difficulty pair (see section A.2 on page 134 of Annex for results). The effect size estimate (Cohen's d_z) was calculated using the t-value (t) divided by square root of the number of observed pairs (n) (Rosenthal, 1991 as referred to by Lakens, 2013, formula 7, p. 4):

$$d_z = \frac{t}{\sqrt{n}} \tag{3.3}$$

All values were entered into the G*Power interface and the resulting computed sample sizes can be found in Table 3.5.

Discussion

The TDRT was expected to require the least amount of participants, followed by the HDRT and RDRT. The HDRT required the least number of participants to differentiate between different levels of load. The number of required participants was inversely related to level of α. Also, baseline differences required more participants than needed in the differentiation of the n-back difficulty levels. For all DRTs in both static and dynamic settings, the primarily visual-manual SuRT would require an exorbitant number of participants to distinguish its difficulty levels as this task did not differ much according to the DRT. In terms of the purely cognitive n-back task, the TDRT and HDRT had similar requirements in terms of measurement points needed to distinguish between the two difficulty levels across static and dynamic conditions. For the RDRT, however, this was not the case and the static RDRT was less sensitive relative to the dynamic condition and therefore required additional participants. For future experiments where a repeated-measures design is used and cognitive tasks of different difficulty levels are evaluated for at least 1-minute, the following recommendations are made (unless otherwise specified or another constraint exists): experiments using the TDRT should test at least 14 persons; the RDRT, at least 19 persons; and finally, the HDRT, at least 8 persons. These values are meant to serve as an approximation of the average number of participants minimally needed to differentiate between a cognitive task load levels in a DRT experiment. They have been determined by taking the higher value of each n-back static/dynamic comparison for each DRT provided Table 3.5, $\alpha = .05$. Increasing the sample size beyond these minimum values is not only highly suggested but will advantageously increase the power of the experimental results.

3.4. Summary and conclusions

An experiment was implemented to investigate the sensitivity of the HDRT, RDRT, and TDRT to variations of task load. Additional POC and power analyses were carried out to investigate whether DRT variants differed in terms of how they time-share with driving and the number of samples needed to reach statistical significance between ST difficulty levels. In the sensitivity assessment (section 3.3.1), DRTs, especially the HDRT, were generally found to vary as a function of increasing task load. This effect was most clear for the n-back and counting tasks under the HDRT, with and without the additional driving task. The SuRT, which varies visual-manual load rather than pure cognitive load, did not produce DRT performance differences. Based on these findings, it is supported that the DRT is most appropriately used in situations where cognitive load is to be measured. DRT performance detriments that were observed for the hard n-back and counting tasks show how test persons attend differently and adapt their reactions to presented signals in order to cope with the increased cognitive demand. As a result of this increased demand, incoming signals are processed (or not processed as suggested by Wickens' model) differently, manifested as additional time needed to respond (i.e., longer RTs) or the lack of responses (miss). Additionally, POC and power analyses revealed the HDRT and TDRT as having the least degree of intrusiveness on other tasks performed concurrently and as requiring the smallest sample sizes. In sum, it can be concluded that in this experiment, the HDRT provided the strongest measure of all the DRTs, followed by the TDRT, and, finally, the RDRT. When deciding on which DRT to implement in an experiment, other issues such as methodological issues (e.g., experimental constraints) should be considered as well as the aforementioned results.

Chapter 4

Experiment II on the reliability and validity of the DRT: EEG & the TDRT

4.1. Overview

In this experiment[1], the TDRT was electrophysiologically assessed in an effort to validate it as a measure of cognitive workload in terms of cognitive state, resulting from task and person based factors. To date, a HDRT-like go/no-go task has been investigated electrophysiologically through an event-related P300 component, where the different task loads were reflected in RTs as well as in a reduction of the P300 component amplitude for static conditions (Strayer et al., 2014 as presented section 2.6.1). However, since DRT experiments often evaluate continuous tasks, participants are assumed to be in a loaded state during these experimental segments. As the P300 component is likely, therefore, related to how task demands affect selective processing and attention, whether the DRT also reflects the general state of being cognitively loaded (i.e., high level of workload, attentive processing) requires further evaluation. To this end, PSDs were used to assess the brain state of participants during various experimental conditions of vary-

---

[1] This experiment was performed in cooperation with an EEG specialist, Fernando Cross Villasana (`http://www.gsn.uni-muenchen.de/people/students/cross_villasana_fernando/index.html` from the Faculty of Sports Psychology, TUM, who provided the EEG equipment, methodological insight and preprocessed the EEG data. All other tasks were performed autonomously by the author of this thesis.

ing task load. Two frequency bands (discussed in section 2.4.1) were relevant in the current experiment: theta and alpha, related to task difficulty and memory load as well as attentional resource allocation, respectively. By evaluating the differences in emerging frequencies for different task load conditions, participants' cognitive states in terms of degree of attention and workload can be identified and measured, which can help to interpret the relationship between cognitive workload as a state and the DRT.

The main research questions addressed in this experiment are:

1. Are the DRT findings replicable (and therefore, reliable)?
2. On the validity of the DRT measure: Are the differences found in DRT performance also reflected in the spectral power of different load conditions? (i.e., does the DRT reflect a state of being cognitively loaded?).

For ease, task conditions might be referred to using abbreviations describing the tasks performed in the condition (e.g., TDN or TDBE). The first letter indicates the DRT used ("T" in this case, for "TDRT"); the second letter indicates whether the simulated driving task was also performed ("D") or not ("N"); and the third letter indicates the ST performed: "N" for none, "B" for n-back, "V" for SuRT; finally, where applicable, the last letter indicates the difficulty of the ST: "E" for easy and "H" for hard.

4.1.1. Hypotheses

The following hypotheses were made pertaining to the research questions identified on the current page:

1. Previous DRT findings were expected to be replicated (relevant to the overall thesis aims provided on page 4). Specifically, the TDRT RTs were expected to be more sensitive than HRs to different task loads in accordance with previous findings (T. A. Ranney et al., 2014; section 3.3.1 of this thesis). Additionally, as previously reported (Engström, 2010; Engström, Larsson, & Larsson, 2013; Hsieh et al., 2015; R. A. Young et al., 2013; section 3.3.1 of this thesis), TDRT RTs were expected to be longer for triple task conditions relative to

baseline as well as during the performance of the hard n-back compared to RTs during concurrent easy n-back task performance. As found in the previous experiment, no differences between the two SuRT difficulty levels were expected. Even if slight experimental differences were present, the replication should at least be present in terms of orders of RT means per condition (as per ISO 17488:2016, Annex E).

2. Alpha and theta amplitudes were expected to decrease and increase, respectively, according to and reflecting a cognitively loaded state and the additional deployment of resources to deal with the task at hand (as per literature reviewed in section 2.4.1). Since different levels of working memory were manipulated in the n-back task, these difficultly levels were expected to be reflected and distinguished in the DRT RT and EEG metrics as per the experiment reported in section 3.3.1 of this thesis for RTs, and Gevins (1997) and Lei and Roetting (2011) for alpha and theta. In contrast, TDRT RTs and EEG metrics were not expected to differ between SuRT easy and hard conditions. Based on these hypotheses and the expected similarity between behavioral and electrophysiological measures, a correlation between the two measures was expected: as RTs rise, alpha levels were expected to decrease and theta levels were expected to increase.

4.2. Methods

4.2.1. Design

A repeated measures design was implemented. Task load was manipulated through different secondary tasks in two levels of difficulty: n-back (a system-paced, primarily cognitively loading task; as reported on page 55) and the SuRT (a user-paced, primarily visual-manual task, as reported on page 55). Due to the typical block experimental design of DRT experiments, spectral analysis was used to assess the continuous demand of the performed tasks. The experiment was separated into 2 blocks: a baseline block and an experimental block. During the experimental block, each level of the n-back and SuRT was tested separately, together with the TDRT

Figure 4.1. – Experimental setup; participant with EEG cap and active electrodes inside the vehicle mock-up.

and a simulated driving task. Baseline performances of the n-back and SuRT tasks, EEG (eyes open and eyes closed), DRT and driving were additionally recorded. Thirteen conditions in total were recorded per participant. All conditions were performed for 2 minutes each in order to have enough data for the EEG. Half of the participants performed the baseline block first and the other half began with the experimental block. Within each block, task order was randomly assigned.

4.2.2. Participants

A total of 20 licensed drivers were tested, 8 of which were excluded: 4 due to excessive EEG artifacts and poor data quality[2], 4 due to technical errors during the recording phase. As a result, the data of 12 participants[3] (N = 12; 7 male and 5 female; 9 right handed and 3 left handed) with an mean age of 26.17 (*SD* = 4.06) qualified for analysis. One participant reported having a red-green color vision impairment but was not disqualified as no effect on task performance was observed. Half of the participants required visual aids (e.g., contact lenses or eyeglasses). On average, participants had possessed their licenses for approximately *M* = 8.17 years (*SD* =

---

[2] The data of 2 participants was not usable due to excessive artifact elimination; for 1 participant, the individual alpha range could not be determined; 1 participant, showed remarkably low alpha levels during baseline, which increased during task engagement.

[3] Sample size was considered adequate as the recording time was doubled in this experiment relative to Experiment I.

Figure 4.2. – Experimental setup inside vehicle mock-up. The monitor placed in the console was used to display the SuRT. SuRT manual inputs were communicated through the numeric keypad located to the left of the gear shift. Near the windshield, 2 cameras (not relevant for this experiment) and an LED (used for EEG calibration) can be seen attached to the windshield via suction cup. The suction cup was removed after the calibration was performed prior to experimentation.

3.23). Eight participants reported to drive more than 5000 km per year; 4 reported driving 5000 km or less per year. All participants were either students or researchers employed at the Chair of Ergonomics, TUM, and were either financially compensated for their participation or volunteered, respectively.

4.2.3. Apparatus

This experiment was carried out in a large static simulator at the Chair of Ergonomics at the TUM. In this simulator, a real, fixed vehicle (BMW 6 series cabriolet; see Figure 4.1) was surrounded by a six-channel projection and sound system[4]. Inside the vehicle and to the right of the driver, a 14-inch flat panel monitor (Lenovo Thinkvision LT1421), used for the SuRT, was located on the center stack (see Figure 4.2) along with a numeric keypad (as per Experiment I). The EEG software ran in parallel as a stand-alone program and was manually triggered. EEG recording was performed with an actiCAP 64 Ag/AgCl active electrode system (Brain Products, Munich, Germany[5]). N-back and TDRT apparatus were used as per Experiment I.

---

[4] For more details, visit: `http://www.lfe.mw.tum.de/en/research/methods-and-lab-equipment/static-driving-simulator/`

[5] http://www.brainproducts.com/

TDRT

The TDRT was implemented in this experiment. A vibration signal was presented every 3 – 5 seconds (OnOn). Each participant experienced 24 – 40 signals per condition. Participants were instructed to respond as quickly and accurately as possible to these signals via button press.

EEG

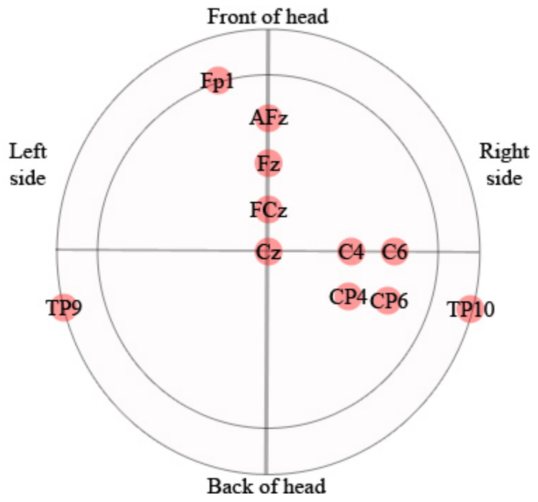


Figure 4.3. – Positions of the electrodes of interest (red circles) used in this experiment; based on the 10-10 system of electrode positioning from American Electroencephalographic Society (1994, p. 112).

EEG was recorded during all conditions. An elastic cap (Easy Cap, FMS) was used to position electrodes, according to the International 10-10 system (American Electroencephalographic Society, 1994; Figure 4.3) on participants' heads. Frontal medial (FM) alpha and theta were assessed using the Fz, FCz, and Cz electrodes, as per Sauseng et al. (2007). To capture effects associated with the placement of the TDRT vibration motor, right sensory motor (RSM) alpha was evaluated through the electrodes: C4, C6, Cp4, and Cp6. These electrodes were determined based on an assessment of the TDRT baseline activity. Similar methods were used as those reported in Cross-Villasana, Gröpel, Doppelmayr, and Beckmann (2015): Online recordings referenced FCz and offline analysis re-referenced to linked (averaged) mastoids (TP9, TP10); AFz served as the ground electrode; a "[v]ertical electrooculogram (VEOG) was registered from an electrode beneath the left eye and Fp1"; "[d]ata was recorded with a Brain Amp amplifier (Brain Products, Munich, Germany[6])", filtered in real time with a "band-pass filter from 0.1 to 250 Hz and a notch filter at 50 Hz, with a sampling rate of 1000 Hz"; and "electrode impedances were kept under 3 kΩ" (Cross-Villasana et al., 2015, p. 4; footnote added). Alpha at RSM and FM regions as well as FM theta were analyzed. The range of each band was individualized by contrasting amplitude values of the alpha and theta bands during periods of 2 minutes with open and closed eyes before the experiment, as described by Klimesch (1999).

Secondary tasks

The n-back task and SuRT were used in this experiment (see page 55 for a detailed explanation).

Simulated driving task

The simulated driving task was identical to that explained in section 3.2.3 on page 54. However, in this experiment performance data were recorded at 300 Hz.

4.2.4. Procedure

The experiment lasted approximately 2 hours. Participants were to arrive at the lab well rested and without having consumed caffeine or nicotine within 2 hours prior to the experiment. A demographic questionnaire was filled out before the training session began. The participants watched a standardized multimedia presentation where each task was explained and instructed. After a task was presented, participants practiced this task until they reported feeling comfortable with it. When all tasks had been trained, participants trained concurrently performing multiple tasks until they reported feeling comfortable and the experimenter was satisfied with their performance. After the training session, the participants were prepared with the EEG equipment prior to the experimental recording. Participants were told which tasks to perform prior to beginning a condition and were also at this time given the opportunity to adjust their lane position and speed as per task instruction.

---

[6] `http://www.brainproducts.com/`

4.2.5. Data preparation

Driving data are provided as RMSE of speed (calculated as per Equation 3.2 on page 71) and standard deviation of lane position (SDLP). SDLP was calculated using the difference between a participant's lane position (*LanePosition*) and the average position (*AveragePosition*) and the count of values per condition (*n*):

$$SDLP = \sqrt{\frac{\sum_{i=1}^{n} \left(LanePosition_i - AveragePosition\right)^2}{n}} \tag{4.1}$$

The DRT and EEG data were the focus of this analysis. The DRT data was checked for cheating to ensure participants were not constantly pressing the response button without being prompted by a signal. To do this, the signal to response ratio was checked. This ratio was under 2[7] and was, therefore, considered acceptable. Only hits and misses were analyzed as the main metrics. The HR[8] was calculated as follows:

$$HR = \frac{\#\, Hits}{\#\, Signals - \#\, Signals\, with\, premature\, responses} \tag{4.2}$$

Where *premature responses* were defined as occurring between 0 - 100 ms post-signal onset. The RTs of hits were were used to compute overall RT means for each participant and condition.

EEG data were prepared similarly to that reported in Cross-Villasana et al. (2015). EEG recordings were segmented according to each recording phase and each of the phases was further segmented into 2 second epochs. A further automatic artifact rejection procedure was implemented for each epoch, using a "maximum allowed amplitude of ±100 µV, maximum allowed voltage steps of 50 µV between consecutive sampling points, and a minimum required signal change of 0.5 µV in 500 ms" (Cross-Villasana et al., 2015, p. 4). Amplitude spectra were calculated using a Fourier Transform with a "Hamming Window with a 50% overlap between contiguous epochs" (Cross-Villasana et al., 2015, p. 4). After the Fourier transform, all epochs were averaged back together. Individual ranges for of the alpha and theta band were calculated for

---

[7] As per the cutoff limit defined in ISO 17488:2016, section 6.10 - Checking data quality.

[8] As defined in ISO 17488:2016, section 3 - Terms and definitions.

each subject according to the method by Klimesch (1999). Before the analysis, the data were visually inspected and all segments containing evident artifacts were rejected. The data was resampled to a power of two (1024 Hz) and filtered using a bandpass infinite-impulse-response (IIR) filter between .05 Hz and 40 Hz. Blink and eye movement artifacts were subtracted using an infomax independent component analysis (ICA).

4.3. Results

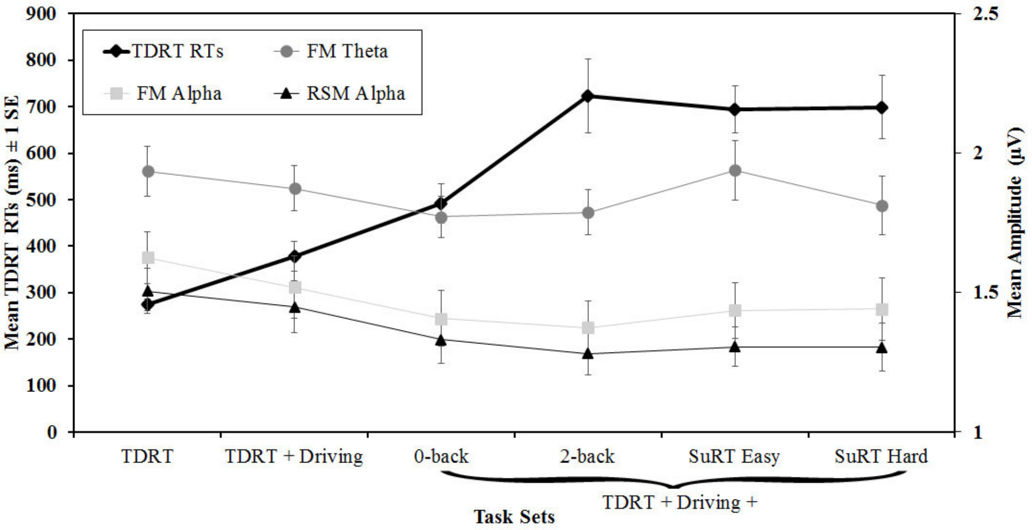


Figure 4.4. – DRT and EEG metrics (± 1 SE) as a function of task set.

Mean RTs as well as alpha and theta amplitudes are plotted in Figure 4.4. Mean DRT RTs of hits and alpha and theta amplitude levels can be found in Table 4.1 on the following page. Median HRs are given in Table 4.3 on page 90. As can be seen in Figure 4.4, longer DRT RTs can be observed for triple task scenarios relative to baselines as well as for the hard n-back condition relative to the easy n-back condition. Mean alpha amplitudes decrease in triple task scenarios relative to baselines. Additionally, higher mean theta amplitudes can be observed for SuRT easy relative to SuRT hard. In order to better assess DRT RTs, the mean RTs of each signal across the task duration is plotted in Figure 4.5 on page 88. As can be seen in Figure 4.5, the mean RTs in both baseline conditions vary less than the mean RTs of other more complex conditions. As the

	Baselines			TDRT + Driving +			
	Driving	TDRT	TDRT + Driving	0-Back	2-Back	SuRT Easy	SuRT Hard
Metric		*M*(*SD*)			*M*(*SD*)		
DRT RT	–	274.38 (42.58)	377.91 (113.80)	492.10 (148.83)	722.56 (274.75)	693.99 (173.67)	699.25 (236.18)
FM theta	–	1.93 (.31)	1.87 (.28)	1.77 (.25)	1.79 (.28)	1.94 (.37)	1.81 (.36)
FM alpha	–	1.63 (.32)	1.52 (.38)	1.41 (.38)	1.38 (.32)	1.44 (.35)	1.44 (.39)
RSM alpha	–	1.51 (.28)	1.45 (.32)	1.33 (.29)	1.28 (.27)	1.31 (.25)	1.30 (.30)
Driving - SDLP	.15 (.05)	–	.14 (.04)	.13 (.04)	.12 (.03)	.25 (.11)	.30 (.11)
Driving - RMSE	3.76 (1.30)	–	3.98 (1.07)	4.50 (1.40)	6.61 (4.04)	7.97 (3.64)	7.59 (2.97)

Table 4.1. – Mean performance values per condition: DRT (RTs [ms]); EEG values (wave amplitudes [µV]; driving (SDLP [m] and RMSE of speed [km/h]).

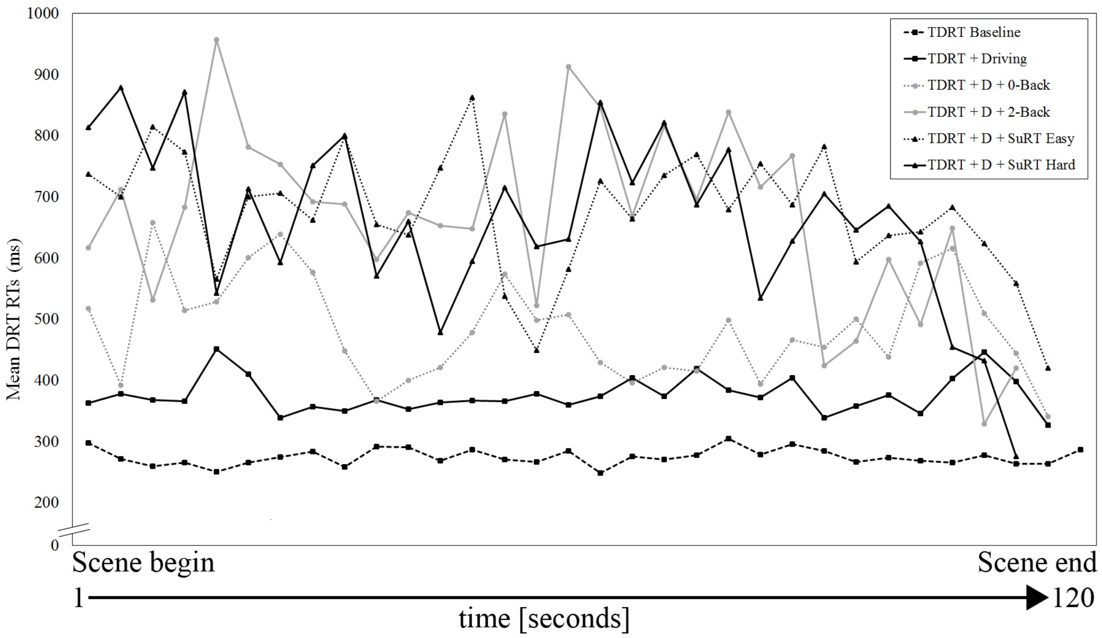


Figure 4.5. – Mean RTs per DRT signal over 2 minute scene duration for all participants (N = 12). Lines represent different task sets.

condition becomes more complex, in the triple task scenarios for example, the mean RTs also become more variable.

Driving SDLP and RMSE values are provided in Table 4.1 on the previous page, and n-back and SuRT performances are reported in Table 4.2; these data were not, however, used for hypothesis testing. Differences in the performance of these tasks were, however, observed. SDLP and RMSE remained similar across the two baselines. No large change between the SDLP during the easy and hard n-back conditions were observed and only a slightly higher SDLP was found for the SuRT hard condition relative to easy. RMSE increased for the triple task scenarios; this

	Baselines		TDRT + Driving +			
	Easy	Hard	0-Back	2-Back	SuRT Easy	SuRT Hard
Task	*M*(*SD*)		*M*(*SD*)			
n-back	0 (0)	1.58 (3.78)	0 (0)	12.58 (10.56)	–	–
SuRT	142.42 (14.77)	42.58 (10.18)	–	–	85.58 (23.08)	24.25 (7.14)

Table 4.2. – Mean performance values for secondary tasks per condition: n-back (error percentage [%]) and SuRT (count of correctly solved trials).

value was higher for the difficult n-back task set, relative to easy, and little difference in the RMSE values between SuRT easy and hard was found. No errors occurred in the performance of the easy n-back task. Errors, however, occurred in the performance of the hard n-back task, which increased when performed together with the TDRT and driving task, relative to baseline. For SuRT, the count of correctly solved trials recorded for baseline trials was higher than those performed together with the TDRT and driving task. Additionally, the easy variant yielded more correctly solved trials than the difficult task.

In the following sections, performance analyses are reported. Parametric analyses were performed despite normality violations as the ANOVA is relatively robust against such violations (see Schmider, Ziegler, Danay, Beyer, & Bühner, 2010). Four 6-way repeated measures ANOVA tests were run to evaluate whether mean DRT RTs, RSM and FM alpha, and FM theta amplitudes were sensitive to differences in task load. One additional non-parametric Friedman test was used to assess the effect of task load on DRT HR distributions. Significant pairwise comparisons of interest, i.e., that between the easy and difficult levels of a ST, are elaborated with mean differences and 95% CI. Pairwise comparisons between STs and baseline measures (DRT for static conditions, DRT with simulated driving for dynamic conditions), are also reported. Pearson correlations were planned to determine the strength of the relationship between the DRT and physiological metrics. However, a visual inspection of the data revealed no linear relationship between the variables and, therefore, correlations were not applicable.

4.3.1. DRT

RTs

Outliers were assessed by box plots. In the TDN and TDBE conditions, the values of 1 participant were higher than others and resulted in outliers. As these values were considered reasonable (685.52 ms for TDN and 866.46 ms for TDBE), they were not omitted from the analysis. Normality was also assessed. According to the Shapiro-Wilk test of normality, the data of all conditions, bar TDN ($p = .007$) and TDBE ($p = .049$), were normally distributed, $p > .05$. Mauchly´s test indicated that the assumption of sphericity had been violated, $\chi^2(14) = 38.17,\ p = .001$, therefore, the Greenhouse-Geisser correction ($\varepsilon = .45$) was used. There was a main effect of task load, $F(2.27, 24.94) = 35.13,\ p < .001$, partial $\eta^2 = .76$. Post-hoc pairwise comparisons with Bonferroni alpha adjustments indicated that static and dynamic baselines were significantly different than all other conditions, $p \leq .01$. The static baseline yielded the fastest RTs, while the dynamic baseline had the second fastest RTs. In terms of differences in difficulty pairs, the n-back hard condition had significantly slower RTs than the easy condition with a mean difference of 230.47 (95% CI [53.16 to 407.78]) ms, $p = .008$. No difference for the SuRT condition difficulty pair was found.

Hit Rates

Condition	Median HR (%)	Minimum HR (%)	Maximum HR (%)
TDRT Baseline	100	100	100
TDRT + Driving	100	100	100
0-Back	100	90.32	100
2-Back	88.49	56.57	100
SuRT easy	100	85.71	100
SuRT hard	95.89	70.00	100

Table 4.3. – TDRT HRs reported as: *Mdn*, minimum and maximum.

A Friedman test was run to determine if there were differences in TDRT HR distributions according to task load. Pairwise comparisons were performed (SPSS 22) with a Bonferroni correction for multiple comparisons. TDRT HRs were significantly different according to task load, $\chi^2(5) = 36.28,\ p < .001$. Post hoc analysis revealed statistically significant differences in TDRT HR between TDBH (*Mdn* = .88) and TDBE (*Mdn* = 1.00), $p = .019$, TDBH and TNN (*Mdn* = 1.00), $p = .004$, TDBH and TDN (*Mdn* = 1.00), $p = .004$. All other comparisons were not statistically significant.

4.3.2. EEG Data

The data of two participants were identified as outliers (see Appendix B.1 on page 137 for box plots), however, were not excluded from the analysis as the values were not judged as extreme. Baseline comparisons were also evaluated and can be found in Appendix B.2 on page 138.

Right Sensory Motor Alpha

Normality was assessed with the Shapiro-Wilk test: 2 conditions (TNN & TDVH) were normally distributed, $p > .05$. The other 4 conditions were not normally distributed: TDN, TDBE, TDBH, and TDVE, $p < .02$. Mauchly′s test indicated that the assumption of sphericity had not been violated, $\chi^2(14) = 19.32,\ p = .166$. There was a main effect of task load, $F(5, 55) = 8.39,\ p < .001,$ partial $\eta^2 = .43$. Post-hoc pairwise comparisons with Bonferroni alpha adjustments indicated that RSM alpha in the TNN condition was significantly different than TDBE, TDBH, and TDVE, $p < .05$. Additionally, the TDN baseline RSM alpha was significantly different than TDBH, $p = .044$. All other comparisons were not significantly different.

Frontal Medial Alpha

Normality was assessed with the Shapiro-Wilk test: 2 conditions (TNN & TDBH) were normally distributed, $p > .05$. The other 4 conditions were not normally distributed: TDN, TDBE, TDVE, and TDVH, $p < .05$. Mauchly′s test indicated that the assumption of sphericity had been violated, $\chi^2(14) = 24.93,\ p = .041$, therefore, the Greenhouse-Geisser correction ($\varepsilon = .54$) was used. There was a main effect of task load, $F(2.72, 29.92) = 10.95,\ p < .001,$ partial $\eta^2 = .50$. Post-hoc pairwise comparisons with Bonferroni alpha adjustments indicated that TNN baseline FM alpha was significantly different than TDBE, TDBH, and TDVE, $p < .02$. The dynamic baseline FM alpha was significantly different than that of TDBE and TDBH, $p < .05$. All other comparisons were not significantly different.

Frontal Medial Theta

Outliers were assessed by box plots. Normality was assessed with the Shapiro-Wilk test: all conditions except 1, TDN, $p = .017$, were normally distributed, $p > .05$. Mauchly´s test indicated that the assumption of sphericity had not been violated, $\chi^2(14) = 15.49,\ p = .362$. There was a main effect of task load, $F(5, 55) = 2.75,\ p = .027,$ partial $\eta^2 = .20$. No post-hoc pairwise comparisons with Bonferroni alpha adjustments resulted as significant.

4.4. Discussion

The aim of this experiment was to evaluate the cognitive state of participants during various experimental conditions of varying task load. EEG activity was recorded during a DRT experiment, and alpha and theta wave amplitudes were assessed. The subsequent sections focus on the two main research questions addressed in this experiment followed by a general discussion.

4.4.1. DRT results are reliable

In the first hypothesis identified in section 4.1.1, it was expected that the previous DRT findings would be replicated and the TDRT RTs would be more sensitive than hits to different task loads. It was also expected that multiple task conditions would yield slower RTs than baseline measures. Reflecting differences in cognitive load, the DRT RTs during the n-back task were expected to be slower for the hard variant and faster for the easier variant. In contrast, no difference in TDRT RTs during SuRT were expected. As in Experiment I, and in accordance with previous research (ex. T. A. Ranney et al., 2014), the RTs were more sensitive than the HR to additional task load. Generally, more significant results were yielded for RTs relative to HR. In this experiment, both the TDRT RTs and HRs significantly differed between the easy and hard n-back tasks and mean TDRT RTs were significantly shorter for baseline conditions than those with an additional secondary task. The secondary tasks implemented differed in difficulty in distinct ways: for the n-back, which is a system-paced task, cognitive demand was varied and required more from working memory as task difficulty increased; for the SuRT, which is a user-paced task, visual-manual demand was varied and required increased visual search and manual input for the harder condition. Only the former difference was reflected in TDRT performance, supporting previous assertions that the TDRT is specific to cognitive demand (Hsieh et al., 2015; R. A. Young et al., 2013).

Although the TDRT results in this study generally replicated those reported in the first experiment, they deviate in terms of statistical significance. Whereas in this experiment task load significantly affected mean RTs in terms of baseline differences and those between the easy and hard n-back task, in the first experiment, the TDRT RTs were significantly affected by task load but all post-hoc tests emerged as not significant. Possible reasons for this lack of statistical significance observed in the previous experiment were already discussed on page 70. Additionally, the increase in recording time used in this experiment could have also been related to the increased sensitivity of the TDRT seen in this experiment. Here, the results yielded by the TDRT resemble the most sensitive DRT reported in the first experiment, the HDRT, and was sensitive to task load differences according to their assumed load level. Based on the many other experiments that have successfully replicated similar findings that the DRT is sensitive to cognitive load (indicated in section 2.6: Conti et al. 2014, 2013, 2012; Engström, Larsson, and Larsson 2013; Harbluk et al. 2013; T. A. Ranney et al. 2014; Siam et al. 2014; R. A. Young et al. 2013) the measure of the DRT is considered reliable.

4.4.2. The DRT reflects attention decrements

The electrophysiological data revealed different aspects of attentive and cognitive processing during DRT performance. Theta and alpha amplitudes were expected to increase and decrease, respectively, as a function of increasing task load (i.e., multiple task conditions relative to baselines), especially for those tasks demanding more working memory load and not for visual-manual load differences. In other words, task conditions involving the n-back hard were expected to have further suppressed alpha amplitudes and greater theta amplitudes relative to n-back easy conditions. Additionally, no amplitude differences were expected between SuRT easy and hard conditions. The alpha amplitudes measured in both FM and RSM regions were significantly affected by task load and found to generally decrease in triple task conditions relative to baselines. Graphically, it could also be observed that the theta amplitude for the SuRT easy variant condition was higher than for the hard variant condition. However, despite overall significant main effects of task load for each physiological measure, these expected task difficulty level differences were not statistically significant in post-hoc tests.

Generally, EEG measures revealed that all triple task conditions require a great degree of attention and processing, possibly indicative of "performance protection" (Hockey, 1997 as referenced by de Waard & Lewis-Evans, 2014, p. 304), in addition to task specific load. This revelation is different than that indicated by the DRT: the DRT is sensitive to and statistically differentiated between expected cognitive load differences. The current findings support those reported by Strayer et al. (2014) and Jahn et al. (2005). Strayer et al. (2014) reported no difference in P300 component amplitude or latency when performing different tasks concurrently with a simulated driving task, despite a significant main effect on P300 amplitude for static conditions. In both Strayer et al. (2014) and Jahn et al. (2005) no difference in heart rate measures were found in conditions of varying difficulty, despite having been sensed through other measures. As suggested by Jahn et al. (2005), such physiological measures can be less selective to the sole effects of a specific construct (e.g., cognitive load) and sensitive to other factors, such as emotional arousal, as well. In this case, the EEG measures were likely more sensitive to overall workload, perhaps including mental effort, which was relatively stable throughout the different task conditions. As noted by O'Donell and Eggemeier in 1986, "[s]econdary task measures are typically considered highly diagnostic and therefore provide an index of load imposed on specific resources. Physiological measures... can either be global (e.g., pupil diameter) or highly diagnostic (e.g., event-related potentials)" (p. 42-4); despite advances in methodology since then, this EEG experiment was not able to support otherwise. As will be discussed in section 4.4.3, the lack of differences in the EEG data could be related to the differing task performance not only observed in the DRT, but in all other tasks. The DRT, however, is sensitive to the way different task demands affect attention selection, which is diagnostic of an effect on attentive mechanisms rather than a measure of global cognitive busyness.

The DRT reflects an event-related response

Earlier, the term cognitive workload was examined, understood to be a general state of *cognitive busyness*. An experiment by Strayer et al. (2014) was also earlier presented where a DRT-like task with the P300 ERP component of EEG was investigated and found to change in amplitude as a function of task load for static conditions. As this component is known to be related to processing time and intensity, it is understood as an indicator of the selection of a stimulus for processing. Through alpha and theta, the relationship between the DRT and measures associated with a cognitively loaded state was evaluated. The wave analysis did not differ between the different task loads, unlike the DRT measures and P300 amplitudes previously reported. Therefore, the measure of the DRT is more specific to a momentary or event related ability to respond to a presented signal in the environment, rather than a pure and sole reflection of a cognitively loaded state. As mentioned earlier, de Waard and Lewis-Evans (2014) (quote on page 24 of this thesis) remark that measurement dissociations detail task performance according to different perspectives, from which a more holistic understanding of task performance may be had. In accordance with this notion, the measurements of both the EEG and DRT are, however, related in that they both reflect different aspects of the same occurrence.

4.4.3. General discussion

In addition to the results discussed in the previous sections, a look at additional performance metrics reveals that these values also varied according to task load. First, RT variability increases when additional tasks are given to perform (see Figure 4.5 and *SD* values reported in Table 4.1), especially when compared to baseline TDRT performance. Moreover, conditions with the hard secondary task variant were also more variable, in terms of RT, compared to conditions with the easy variant. The fluctuations in RT themselves appear random and indicate a task switching strategy adopted by participants to cope with increased task demands. Indeed, this observation may also show a less conscious processing strategy taken on by the cognitive system. In accordance with the current findings, in order to perform many complex tasks at once, serial task engagement may occur in order to be able to maintain performance of each individual task as needed. This finding supports the concept of a central processing limitation, where due to the high level of task demand, participants serially allocate attention and effort to each task (similar to that discussed in Botvinick & Cohen, 2014). Engström (2010) also drew a similar conclusion based on the lack of major difference between DRT modalities. He states that the measure of this method "...relates to some kind of higher-level, amodal, 'cognitive interference in attention selection" (p. 98). This idea of serial task engagement due to processing, control, or resource limitations is compatible with the models of attention presented in section 2.2.

Similarly, the RT variance observed in the different conditions (see Figure 4.5 on page 88) suggests that complex task performance requirements are dealt with through a variable, serial mechanism. A supporting explanation is offered by the controlled parallel processing model of Pashler (1998a), whereby a condition's load not exceeding a processing threshold (e.g., baseline conditions) would be able to be processed and performed in parallel. As soon as this threshold is exceeded (e.g., triple task conditions), serial processing occurs. As evident in mean RTs, not only does the variability increase with demand, RTs increase, oscillating at a higher level according to task load. Best depicted by the n-back task difficulty pair, the bandwidth within which DRT reactions are given for the n-back hard condition is offset to a higher range relative to the easy condition, illustrative of the increasing load of the hard condition.

Not only was TDRT performance affected by task load, performance decrements were observed in all tasks (see Tables 4.1 on page 87 and 4.2 on page 88). The median TDRT hit rate, for example, reached 100% during baseline measurements, but dropped as soon as a secondary task was added, and even further when the secondary task was difficult. Relative to baseline, performance of the easy and hard SuRT decreased approximately 40%. Additionally, participants performed on average over 3 times as many correctly solved scenes in the easy condition as in the hard condition; undoubtedly affecting processing mechanisms and brain state.

Interestingly, TDRT HR, rather than the RT, trended towards a decrease during the SuRT hard condition relative to the easy condition, reflecting the individual's inability to attend and respond to the vibrating motor when the SuRT was more difficult to perform. In support of this, although alpha remains steadily activated, possibly at maximum (a possible floor effect), theta amplitude is higher (although not statistically significant) for the easy SuRT condition than for the difficult SuRT. Here, a potential dissociation of the DRT metric can be made: DRT RTs are diagnostic of the attentive effects of cognitive load and the DRT HR, based on its similarity to the theta measure, could be considered as more sensitive to global workload including also workload associated with visual tasks or those requiring additional control (as per e.g., Cavanagh & Frank, 2014; Gevins, 1997). Van der Horst and Martens (2010), for example, noted the sensitivity of DT performance accuracy to simulated driving task difficulty differences. Therefore, HRs, although less sensitive than RTs, certainly contribute to the overall understanding of a participant's cognitive state but qualitatively differently than RTs.

Despite the ability of the DRT to yield differences in task demand, the physiological measures were not as specifically sensitive. A possible explanation for a lack of mean amplitude variation across different task difficulties is that the EEG could have been sensitive to other psychological or physiological effects, confounding and concealing the effects of cognitive load (as suggested by Jahn et al., 2005). It is also possible that participants adopted different performance or control strategies depending on the task difficulty, as discussed earlier and similar to that suggested by Mantzke and Keinath (2015). Additionally, the metrics were analyzed according to blocks rather than event-related, decided on for this experiment based on the use consensus of the DRT: a mean RT is used to represent the load of a task condition provided the mean HR is above 80% (ISO 17488:2016). It is possible that by averaging the EEG signal across the 2 minute scene duration, variations in task load were obscured. Critics of event-related studies often claim that introducing a stimulus or signal to a condition for measurement, artificially increases the workload intended to be measured (see Tsang & Vidulich, 2006, p. 256; Berka et al., 2007, p. B232). In line with this, it is therefore possible that the TDRT served as an additional source of task load (in line with the POC analysis presented in 3.3.2). Similarly, it is possible that the simulated driving task increased load to an extent that concealed secondary task load specific effects (similar to the findings reported in Strayer et al. [2014], where conditions involving a simulated driving task no longer revealed a significant effect of task load on P300 amplitude). Future electrophysiological studies should test (1) secondary tasks with and without the DRT, without the simulated driving task (viz., static), outside of the simulator (to avoid possible simulator-induced biological noise, as noted by Strayer and colleagues [2014]), and (2) with the simulated driving task without the DRT. These measures will enable the evaluation of the extent to which the driving task, simulator, and DRT, respectively, cause additional load and whether physiological measures can better approximate task load under different conditions.

4.5. Conclusions

The DRT results were similar to those previously reported and support the reliability of the DRT and its use in measuring cognitive load differences. Through the presented wave analysis, it was revealed that multitask conditions including the DRT are attentionally and cognitively demanding. The EEG technique used in this experiment reflected cognitive workload as a mental state in an applied setting. The TDRT reflected one's ability to respond to a signal while performing other tasks; a more selective event than the a global reflection of workload. Although the measures dissociated, both EEG and DRT reflected different aspects of workload from two different perspectives. A direct correlation between the behavioral (DRT) and electrophysiological (EEG) data could not be established due to the inconsistent performance of the tasks tested and possible additional sensitivities of the electrophysiological measures to factors other than task load.

Chapter 5

General discussion

The overarching theme of this thesis was to investigate the sensitivity (Experiment I), reliability and validity (Experiment II) of the DRT. In the first experiment, the HDRT, RDRT, and TDRT were evaluated in terms of their sensitivity to different levels and types of task demand. The second experiment was implemented in order to evaluate the reliability of the DRT and its validity as a measure of a cognitively loaded state. The results of the first experiment showed that the HDRT, RDRT and TDRT were sensitive to differing levels of cognitive load: generally, hit rates fell with increasing demand and RTs, the more sensitive metric, rose. Additionally, the HDRT resulted as the most sensitive measure, produced the most statistically significant results, required the least number of test persons and time-shared best with the concurrent driving task. However, in terms of the overall data trends, no major differences between the DRTs were observed (as per Conti et al., 2012; Engström, Johansson, & Östlund, 2005; Merat & Jamson, 2008; T. A. Ranney et al., 2014; Siam et al., 2014), supporting its sensitivity on an "amodal" level (as per Engström, 2010, p. 98). The effect of cognitive task load on DRT RTs was clearest for the counting and n-back tasks and differences between conditions with and without the additional driving task. The SuRT, a visual-manual, user-paced task, did not produce DRT RT performance differences (similar to for ex. R. A. Young et al., 2013). In both experiments, the DRT RTs for the hard n-back task and both SuRT difficulty levels were similar, trending to be higher for the SuRT variants. Since the SuRT is a user-paced task, it requires the coordination of many different processes. Therefore, it could be that this task, despite the manipulated difficulty level, is generally more cognitively demanding in terms of coordination, control and response processes than the tested system-paced n-back task. However, as the pacing of tasks was not specifically manipulated in these experiments, further research would be needed in order to determine whether this is the case.

In the second experiment, DRT trends observed in the first experiment were replicated, supporting other studies that reported similar findings (ISO 17488:2016, Annex E). Even though the general DRT findings were reproduced in the second experiment, a comparison of the TDRT in Experiment I and II reveals an increase in sensitivity in the latter experiment. Although this does not directly challenge the method's reliability, it nevertheless requires interpretation. Indeed different test persons were tested in each experiment, a fact which could have lead to differences in the results of the experiments. In the first experiment, it is possible that due to the increased number of tasks and tested conditions, other effects such as task order, boredom, vigilance, could have influenced the results. It is also possible that the effects in the second experiment, relative to Experiment I, were more clear due to the increased (doubled) test time and a decrease in evaluated tasks as well as overall experimentation time. Although efforts were made to avoid the aforementioned issues in the first experiment, such as condition randomization across participants, all task combinations were not able to be tested.

Validity of the DRT in terms of its direct reflection of a cognitively loaded state was investigated, however could not be supported. Rather, a dissociation between the electrophysiological and behavioral results was observed. This finding reinforced the interpretation of the performance of the DRT as reflecting how one reacts to stimuli, providing the stimulus has been sensed, perceived and can be responded to. The speed at which and whether a reaction is given to the DRT signal details how attention is affected by the concurrent performance of a task of interest. Although the performance of the DRT is related to a cognitively loaded state, depending on task load, effort, etc., the dissociation between the DRT and EEG measures indicates that these two measures are not identical and reflect different aspects of cognitive workload.

In terms of traffic safety, it was established earlier in this thesis that the DRT measure is conditionally related to traffic safety and accident risk (section 2.6.4). Specifically, that loading a driver and measuring a high RT while a given task is performed, does not mean that the driver will be distracted from the driving task and cause an accident. Previous research (e.g., Mantzke & Keinath, 2015) was not able to establish a relationship between DRT metrics and one's ability to respond in a critical situation. It is, however, possible that this type of question need not, and perhaps cannot, be addressed through experimental manipulation, which would require the direct comparison of two distinct conditions: whereas the DRT requires continuous performance of a detection and response sequence, reacting to a sudden, critical event can be automatic (as per attention models of Engström, Victor, and Markkula 2013; Trick and Enns 2009; and Engström, 2011, Paper III). The performance of the DRT reflects task-related changes in attention and is not intended to directly specify one's response to a critical situation. Additionally, responding to the DRT is executed by a quick and discrete button press; braking, for example, in response to a critical event on the road has an intensity in addition to an initiation time. If able to be perceived, a potential critical event serves as a signal that can activate pathways or schemata that can be used to avoid danger. Presumably, the critical events presented to participants in previous studies (Mantzke & Keinath, 2015; Engström, 2011, Paper III) were able to be processed and responded to despite task load.

Regarding driver distraction, the DRT is sensitive to the type of distraction potential cognitive tasks are associated with: object and event detection performance (as per Tijerina, 2000 and Figure 2.6). As such, the DRT delivers the probability of a reaction to a signal and a RT estimation while performing the evaluated task. Taking into consideration the findings of Mantzke and Keinath (2015) and Engström (2011, Paper III), it is possible that these predictions are only able to be limited to reactions to non-critical signals. In order to be able to reduce the distraction (potential) associated with such tasks, the factors leading to multiple task performance as well as strategy/performance failures need to be addressed. As presented earlier, Vollrath et al., 2016 found that in the real-world, although most persons do not perform additional tasks behind the wheel, those that do do so regardless of whether they are driving or under easier conditions such as when stopped at a traffic light. Due to the nature of this study, the authors were not able to observe possible preparatory actions prior to secondary task engagement (see J. Y. Lee et al., 2016), nor were they able to inquire about the external, social and psychological factors related to performing a task at a given moment or not. Both of these specifications, however, would be interesting to detail in order to be able to better understand such behavior. Some authors have suggested that "task experience" (Groeger, 2001, p. 133) and the resulting learned associations are related to how drivers understand, define, and ultimately perceive dangerous and difficult situations (Groeger, 2001, pp. 121-141). Applied to findings of Vollrath et al. (2016), it is possible that the past experiences of individuals, or lack thereof, had some effect on their choice to engage in secondary task performance. If this is true, traffic safety can be increased through specialized training sessions, instructing drivers on the dangers of extraneous task performance behind the wheel as well as how to anticipate and cope with critical events in case they arise. Such sessions could take place on a safe, designated road or even electronically. Petzoldt, Weiß, Franke, Krems, and Bannert (2013), for example, found that drivers who had trained their hazard anticipation skills through a computer-based training, were able to more quickly identify hazards within the simulated scenery compared to control participants and those having completed a paper-based training. By training drivers how to anticipate and cope with critical situations, new schemata would be created that could be employed in case a real situation on the road occurs (as per Engström, Victor, & Markkula, 2013).

Another way to reduce distraction potential is through appropriately designed and tested in-vehicle devices and concepts. The DRT was developed exactly for this purpose and can assist designers in evaluating whether different device interactions or tasks are associated with attentional decrements to a suboptimal degree.

Throughout this dissertation, several possible future directions have been suggested. How DRT performance is affected by visual DRT signal properties or task pacing, for example would address fundamental research questions relevant for result comparisons across studies as well as for in-vehicle task design. EEG studies including conditions and settings additional to those presented in this thesis were also suggested on page 97. In addition to these suggestions, future research should scrutinize whether the predictive value of the performance of the DRT is in fact only limited to non-critical signals. As previously mentioned, proper design can avoid and reduce cognitive distraction. Relevant to this, future efforts should be made in establishing a grade system according to which a task or device may pass or fail depending on the DRT performance associated with it. At the moment, such a system does not exist. Additionally, research can evaluate the effectiveness of risk anticipation and critical event training sessions. If such sessions are successful, they could offer a way to increase traffic safety through education and preventative techniques. Finally, how RT studies like those reported in this work relate to automated driving should also be considered. Damböck (2013) and Gold (expected: 2016), for example, investigated RTs in terms of assessing take-over performance. Although the importance of RT studies have already been recognized by researchers in automated driving (e.g., Damböck and Gold), much work in this field is still needed.

In conclusion, the DRT method can be used to quantify the cognitive demand a task places on a human operator, in terms of his or her ability to attend and respond to external stimuli. Additionally, its results are reliable and can be replicated across experimental settings. Although the DRT is a very sensitive, powerful tool, it is up to the researcher to ensure that conclusions based on resulting metrics and observations do not extend the capability of the research tool used. Based on the experiments presented in this thesis, the following recommendations on how to understand and use the DRT are made:

1. The DRT is sensitive to and measures cognitive task load.
2. If experimental constraints permit, use HDRT or TDRT for best results.
3. DRT results are reliable and can be replicated across studies.
4. DRT results are valid in terms of how selective attention changes while performing a task of interest.
5. Based on the DRT performance, driver distraction potential, in terms of object and event detection, may be inferred. This potential, however, may only be applicable to non-critical signals.

Bibliography

Allport, A. (1993). Attention and control: have we been asking the wrong questions? A critical review of twenty-five years. In D. Meyer & S. Kornblum (Eds.), *Attention and performance xiv, synergies in experimental psychology, artificial intelligence, and cognitive neuroscience* (pp. 183–218). Cambridge: MIT Press.

Almahasneh, H., Chooi, W.-t., Kamel, N., & Saeed, A. (2014). Deep in thought while driving: An EEG study on drivers' cognitive distraction. *Transportation Research Part F*, *26*, 218–226. doi: 10.1016/j.trf.2014.08.001

Amado, S., & Ulupinar, P. (2005). The effects of conversation on attention and peripheral detection: Is talking with a passenger and talking on the cell phone different? *Transportation Research Part F: Traffic Psychology and Behaviour*, *8*(6), 383–395. doi: 10.1016/j.trf.2005.05.001

American Encephalographic Society. (1994). Guideline Thirteen: Guidelines for standard electrode position nomenclature. *Journal of Clinical Neurophysiology*, *11*(1), 111–113.

Angell, L., Auflick, J., Austria, P., Kochhar, D., Tijerina, L., Biever, W., … Kiger, S. (2006). Driver workload metrics project task 2 final report. (DOT HS 810635), National Highway Traffic Safety Administration.

Anguera, J. A., Boccanfuso, J., Rintoul, J. L., Al-Hashimi, O., Faraji, F., Janowich, J., … Gazzaley, A. (2013). Video game training enhances cognitive control in older adults. *Nature*, *501*(7465), 97–101. doi: 10.1038/nature12486

Baddeley, A. D., & Hitch, G. (1974). Working memory. In G. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 8, pp. 47–89). New York:

Academic Press.

Békésy, G. (1957). Neural volleys and the similarity between some sensations produced by tones and by skin vibrations. *The Journal of the Acoustical Society of America*, *29*(10), 1059–1069.

Bengler, K. (2014). Driver Distraction. In D. Crolla, D. E. Forster, T. Kobayashi, & N. Vaughan (Eds.), *Encyclopedia of automotive engineering* (pp. 3575–3582). Chichester, UK: John Wiley & Sons, Ltd. doi: 10.1002/9781118354179.auto189

Bengler, K., Kohlmann, M., & Lange, C. (2012). Assessment of cognitive workload of in-vehicle systems using a visual peripheral and tactile detection task setting. *Work*, *41*(Suppl 1), 4919–4923. doi: 10.3233/WOR-2012-0786-4919

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

Blalock, L. D., Sawyer, B. D., Kiken, A., Gutzwiller, R. S., McGill, C. L., & Clegg, B. A. (2014). Cognitive load while driving impairs memory of moving but not stationary elements within the environment. *Journal of Applied Research in Memory and Cognition*, *3*(2), 95–100. doi: 10.1016/j.jarmac.2014.04.006

Boff, K., & Lincoln, J. (1988a). Factors affecting simple reaction time. In *Engineering data compendium: Human perception and performance* (pp. 1856–1857). Ohio: AAMRL. Wright-Patterson AFB.

Boff, K., & Lincoln, J. (1988b). Performance Operating Characteristic. In *Engineering data compendium: Human perception and performance* (pp. 1410–1411). Ohio: AAMRL. Wright-Patterson AFB.

Boff, K., & Lincoln, J. (1988c). Reaction time tasks and variability. In *Engineering data compendium: Human perception and performance* (pp. 1842–1843). Ohio: AAMRL. Wright-Patterson AFB.

Boff, K., & Lincoln, J. (1988d). Simple reaction time to visual targets. In *Engineering data compendium: Human perception and performance* (pp. 1858–1859). Ohio: AAMRL. Wright-

Patterson AFB.

Boff, K., & Lincoln, J. (1988e). Simple reaction time to visual targets of different luminances. In *Engineering data compendium: Human perception and performance* (pp. 1844–1845). Ohio: AAMRL. Wright-Patterson AFB.

Bolanowski, S. J. J., Gescheider, G., Verrillo, R., & Checkosky, C. (1988). Four channels mediate the mechanical aspects of touch. *The Journal of the Acoustical Society of America*, *84*(5), 1680–1694.

Boles, D. (2001). Multiple Resources. In W. Karwowski (Ed.), *International encyclopedia of ergonomics and human factors* (pp. 271–275). London: Taylor & Francis.

Bolles, R. C. (1993). Sensory physiology and perceptual problems. In *The story of psychology: A thematic history* (pp. 97–120). California: Brooks/Cole Publishing.

Botvinick, M. M., Braver, T. S., Barch, D. M., Carter, C. S., & Cohen, J. D. (2001). Conflict monitoring and cognitive control. *Psychological Review*, *108*(3), 624–652. doi: 10.1037//0033-295X.I08.3.624

Botvinick, M. M., & Cohen, J. D. (2014). The computational and neural basis of cognitive control: Charted territory and new frontiers. *Cognitive Science*, *38*(6), 1249–1285. doi: 10.1111/cogs.12126

Boyce, P. R. (2006). Illumination. In G. Salvendy (Ed.), *Handbook of human factors and ergonomics* (3rd ed., pp. 643–673). New Jersey: John Wiley & Sons.

Broadbent, D. (1958). *Perception and Communication*. Oxford, UK: Pergamon Press.

Brookhuis, K., & de Waard, D. (2001). Assessment of drivers' workload: performance and subjective and physiological indexes. In P. A. Hancock & P. A. Desmond (Eds.), *Stress, workload, and fatigue* (pp. 321–333). New Jersey: Lawrence Erlbaum Associates.

Bruns, A. (2004). Fourier-, Hilbert- and wavelet-based signal analysis: Are they really different approaches? *Journal of Neuroscience Methods*, *137*(2), 321–332. doi: 10.1016/j.jneumeth.2004.03.002

Bruyas, M.-P., & Dumont, L. (2013). Sensitivity of detection response task (DRT) to the driving demand and task difficulty. In *Proceedings of the seventh international driving symposium* *on human factors in driver assessment, training, and vehicle design* (pp. 64–70). Iowa City: University of Iowa. Retrieved from `http://drivingassessment.uiowa.edu/sites/default/files/DA2013/Papers/011_Bruyas_0.pdf`

Bubb, H. (2015). Einführung. In H. Bubb, K. Bengler, R. E. Grünen, & M. Vollrath (Eds.), *Automobilergonomie* (pp. 1–25). Weisbaden: Springer.

Bubb, H., Bengler, K., Lange, C., Aringer, C., Trübswetter, N., Conti, A., & Zimmerman, M. (2015). Messmethoden. In H. Bubb, K. Bengler, R. E. Grünen, & M. Vollrath (Eds.), *Automobilergonomie* (pp. 617–662). Wiesbaden: Springer.

Burns, P. C., Bengler, K., & Weir, D. H. (2010). Driver Metrics, an Overview of User Needs and Uses. In G. L. Rupp (Ed.), *Performance metrics for assessing driver distraction: The quest for improved road safety* (pp. 24–30). Warrendale, PA, USA: SAE International.

Cavanagh, J. F., & Frank, M. J. (2014). Frontal theta as a mechanism for cognitive control. *Trends in Cognitive Sciences*, *18*(8), 414–421. doi: 10.1016/j.tics.2014.04.012

Champely, S., Ekstrom, C., Dalgaard, P., Jill, J., & De Rosario, H. (2015). *Package "pwr": Basic Functions for Power Analysis.* Retrieved from `https://cran.r-project.org/web/packages/pwr/pwr.pdf`

Chelazzi, L., Duncan, J., Miller, E. K., & Desimone, R. (1998). Responses of neurons in inferior temporal cortex during memory- guided visual search. *Journal of Neurophysiology*, *80*(6), 2918–2940.

Cohen, B. H. (2013a). Introduction to hypothesis testing: the one-sample z test. In *Explaining psychological statistics* (4th ed ed., pp. 135–172). New Jersey: John Wiley & Sons.

Cohen, B. H. (2013b). Statistical Power and Effect Size. In *Explaining psychological statistics* (4th ed. ed., pp. 237–270). Hoboken, New Jersey: John Wiley & Sons.

Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*(1), 155–159. doi: 10.1037/0033-2909.112.1.155

Cohen, J. (2009). *Statistical power analysis for the behavioral sciences* (2nd ed.). East Sussex: Psychology Press.

Cohen, J. D., Dunbar, K., & McClelland, J. L. (1990). On the control of automatic processes: A parallel distributed processing account of the stroop effect. *Psychological Review*, *97*(3), 332–361.

Conti, A. S., Dlugosch, C., & Bengler, K. (2014). The effect of task set instruction on detection response task performance. In D. de Waard et al. (Eds.), *Proceedings of the human factors and ergonomics society europe chapter 2013 annual conference* (pp. 107 – 117). Retrieved from `http://www.hfes-europe.org/wp-content/uploads/2014/06/Conti.pdf`

Conti, A. S., Dlugosch, C., Schwarz, F., & Bengler, K. (2013). Driving and speaking: revelations by the head-mounted detection response task. In *Proceedings of the seventh international driving symposium on human factors in driver assessment, training, and vehicle design* (pp. 362–368). Iowa City: University of Iowa. Retrieved from `http://drivingassessment.uiowa.edu/sites/default/files/DA2013/Papers/056_Conti_0.pdf`

Conti, A. S., Dlugosch, C., Vilimek, R., Keinath, A., & Bengler, K. (2012). An assessment of cognitive workload using detection response tasks. In N. A. Stanton (Ed.), *Human aspects of road and rail transportation* (pp. 735–743). CRC Press. Retrieved from `http://www.crcnetbase.com/doi/abs/10.1201/b12320-82`

Conti, A. S., Krause, M., Späth, M., & Bengler, K. (2015). *Different motors, different DRT metrics.* doi: 10.13140/RG.2.1.1307.5920

Cooper, J. M., Medeiros-Ward, N., & Strayer, D. L. (2013). The impact of eye movements and cognitive workload on lateral position variability in driving. *Human Factors*, *55*(5), 1001–1014. doi: 10.1177/0018720813480177

Cross-Villasana, F., Gröpel, P., Doppelmayr, M., & Beckmann, J. (2015). Unilateral left-hand contractions produce widespread depression of cortical activity after their execution. *Plos One*, *10*(12), e0145867. doi: 10.1371/journal.pone.0145867

Daanen, H. A. M., van De Vliert, E., & Huang, X. (2003). Driving performance in cold, warm, and thermoneutral environments. *Applied Ergonomics*, *34*(6), 597–602. doi: 10.1016/S0003-6870(03)00055-3

Damböck, D. (2013). *Automationseffekte im fahrzeug – von der reaktion zur übernahme* (Doctoral thesis). Technical University of Munich, Munich, Germany.

de Waard, D. (1996). *The measurement of drivers' mental workload* (Doctoral thesis). Retrieved from `http://www.rug.nl/research/portal/files/13410300/09_thesis.pdf`

de Waard, D., & Lewis-Evans, B. (2014). Self-report scales alone cannot capture mental workload. *Cognition, Technology & Work*, *16*(3), 303–305. doi: 10.1007/s10111-014-0277-z

Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annual Review of Neuroscience*, *18*, 193–222. doi: 10.1146/annurev.neuro.18.1.193

Deutsch, J. A., & Deutsch, D. (1963). Attention: Some theoretical considerations. *Psychological Review*, *70*, 80–90. doi: 10.1037/h0039515

Diels, C. (2011). Tactile detection task as a real time cognitive workload measure. In P. Bust (Ed.), *Contemporary ergonomics 2011* (pp. 183–190). London: Taylor & Francis.

Dienes, Z. (2008). *Understanding psychology as a science: An introduction to scientific and statistical inference*. Houndmills: Palgrave Macmillian.

Donders, A. R. T., van der Heijden, G. J., Stijnen, T., & Moons, K. G. (2006). Review: A gentle introduction to imputation of missing values. *Journal of Clinical Epidemiology*, *59*, 1087–1091. doi: 10.1016/j.jclinepi.2006.01.014

Donders, F. C. (1969). On the speed of mental processes. *Acta psychologica*, *30*, 412–431. doi: 10.1016/0001-6918(69)90065-1

Engström, J. (2010). The tactile detection task as a method for assessing drivers' cognitive load. In G. L. Rupp (Ed.), *Performance metrics for assessing driver distraction: The quest for improved road safety* (pp. 90 – 103). Warrendale: SAE International.

Engström, J. (2011). *Understanding attention selection in driving* (Doctoral thesis). Chalmers University of Technology, Gothenburg, Sweden.

Engström, J., Åberg, N., Johansson, E., & Hammarbäck, J. (2005). Comparison between visual and tactile signal detection tasks applied to the safety assessment of in-vehicle information systems. In *Proceedings of the third international driving symposium on human factors in driver assessment, training and vehicle design* (pp. 232–239). Iowa City: University of Iowa. Retrieved from `https://trid.trb.org/view.aspx?id=763213`

Engström, J., Johansson, E., & Östlund, J. (2005). Effects of visual and cognitive load in real and simulated motorway driving. *Transportation Research Part F: Traffic Psychology and Behaviour*, *8*, 97–120. doi: 10.1016/j.trf.2005.04.012

Engström, J., Larsson, P., & Larsson, C. (2013). Comparison of static and driving simulator venues for the tactile detection response task. In *Proceedings of the seventh international driving symposium on human factors in driver assessment, training, and vehicle design* (pp. 369–375). Iowa City, USA: University of Iowa. Retrieved from `http://drivingassessment.uiowa.edu/sites/default/files/DA2013/Papers/057_Engstrom_0.pdf`

Engström, J., Victor, T., & Markkula, G. (2013). Attention selection and multitasking in everyday driving: A conceptual model. In M. A. Regan, J. D. Lee, & T. W. Victor (Eds.), *Driver distraction and inattention: Advances in research and countermeasures* (Vol. 1, pp. 27–54). Burlington: Ashgate.

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G * Power 3 : A flexible statistical power analysis program for the social , behavioral , and biomedical sciences. *Behavior research methods*, *39*(2), 175–191. doi: 10.3758/BF03193146

Feng, S. F., Schwemmer, M., Gershman, S. J., & Cohen, J. D. (2014). Multitasking versus multiplexing: Toward a normative account of limitations in the simultaneous execution of control-demanding behaviors. *Cognitive, Affective, & Behavioral Neuroscience*, *14*(1), 129–146. doi: 10.3758/s13415-013-0236-9

Field, A. (2013). *Discovering statistics using IBM SPSS statistics* (4th ed.). London: Sage.

Foley, J., Young, R., Angell, L. A., & Domeyer, J. (2013). Towards operationalizing driver distraction. In *Proceedings of the seventh international driving symposium on human factors in driver assessment, training, and vehicle design* (pp. 57–63). Iowa City, USA: University of Iowa. Retrieved from `http://drivingassessment.uiowa.edu/sites/default/files/DA2013/Papers/010_Foley_0.pdf`

Frey, J. N., Ruhnau, P., Leske, S., Siegel, M., Braun, C., & Weisz, N. (2016). The tactile window to consciousness is characterized by frequency-specific integration and segregation of the primary somatosensory cortex. *Scientific Reports*, *6*. doi: 10.1038/srep20805

Fuchs, A. H., & Milar, K. S. (2003). Psychology as a science. In D. K. Freedheim & I. B. Weiner (Eds.), *Handbook of psychology* (Vol. 1, pp. 1–26). Hoboken, New Jersey: John Wiley & Sons.

Fuller, R. (2005). Towards a general theory of driver behaviour. *Accident Analysis and Prevention*, *37*(3), 461–472. doi: 10.1016/j.aap.2004.11.003

Fuller, R. (2011). Driver control theory: From task difficulty homeostasis to risk allostasis. In B. E. Porter (Ed.), *Handbook of traffic psychology* (pp. 13–26). London: Academic Press. doi: 10.1016/B978-0-12-381984-0.10002-5

Gawron, V. J. (2008). *Human performance, workload, and situational awareness measures handbook* (2nd ed. ed.). Boca Raton, Florida: Taylor & Francis Group.

Gescheider, G., Bolanowski, S., & Verrillo, R. (2004). Some characteristics of tactile channels. *Behavioural Brain Research*, *148*(1-2), 35–40. doi: 10.1016/S0166-4328(03)00177-3

Gevins, A. (1997). High-resolution EEG mapping of cortical activation related to working memory: Effects of task difficulty, type of processing, and practice. *Cerebral Cortex*, *7*(4), 374–385. doi: 10.1093/cercor/7.4.374

Gold, C. (expected: 2016). *Modeling of take-over situations in highly automated vehicle guidance* (Doctoral thesis). Technical University of Munich, Munich, Germany.

Gopher, D., & Donchin, E. (1986). Workload—An Examination of the Concept. In K. R. Boff, L. Kaufman, & J. P. Thomas (Eds.), *Handbook of perception and human performance* (Vol. 2, pp. 41–1 – 41–49). New York: Wiley.

Groeger, J. A. (2001). *Understanding driving: Applying cognitive psychology to a complex everday task* (Reprinted ed.). East Sussex: Psychology Press.

Grondin, S., & Macar, F. (1992). Dividing attention between temporal and nontemporal tasks: a performance operating characteristic -POC- analysis. In F. Macar, V. Pouthas, & W. J. Friedman (Eds.), *Time, action and cognition: Towards bridging the gap* (Vol. 66, pp. 119–128). Netherlands: Kluwer Academic Publishers.

Halata, Z., & Baumann, K. (2008). Anatomy of receptors. In M. Grunwald (Ed.), *Human haptic perception: Basics and applications* (pp. 137–163). Basel: Birkhäuser.

Hancock, P. A., Meshkati, N., & Robertson, M. M. (1985). Physiological reflections of mental workload. *Aviation Space and Environmental Medicine*, *56*(11), 1110–1114.

Hanson, J. V., Whitaker, D., & Heron, J. (2009). Preferential processing of tactile events under conditions of divided attention. *Neuroreport*, *20*(15), 1392–1396.

Harbluk, J. L., Burns, P. C., Hernandez, S., Tam, J., & Glazduri, V. (2013). Detection response tasks: using remote, headmounted and tactile signals to assess cognitive demand while driving. In *Proceedings of the seventh international driving symposium on human factors in driver assessment, training, and vehicle design* (pp. 78–84). Iowa City, USA: University of Iowa. Retrieved from `http://drivingassessment.uiowa.edu/sites/default/files/DA2013/Papers/013_Harbluk_0.pdf`

Harbluk, J. L., Noy, Y. I., Trbovich, P. L., & Eizenman, M. (2007). An on-road assessment of cognitive distraction: Impacts on drivers' visual behavior and braking performance. *Accident Analysis & Prevention*, *39*(2), 372–379. doi: 10.1016/j.aap.2006.08.013

Heitz, R. P. (2014). The speed-accuracy tradeoff: History, physiology, methodology, and behavior. *Frontiers in Neuroscience*, *8*, 1–19. doi: 10.3389/fnins.2014.00150

Ho, C., & Spence, C. (2008). *The multisensory driver: Implications for ergonomic car interface design*. Hampshire: Ashgate.

Hsiao, S., & Yau, J. (2008). Neural basis of haptic perception. In M. Grunwald (Ed.), *Human haptic perception: Basics and applications* (pp. 103–112). Basel: Birkhäuser.

Hsieh, L., Seaman, S., & Young, R. (2015). *A surrogate test for cognitive demand: Tactile detection response task (TDRT)* (Tech. Rep.). Retrieved from `http://papers.sae.org/2015-01-1385/` doi: 10.4271/2015-01-1385

Huffman, K. (2003). *Psychology in Action*. Hoboken, New Jersey: John Wiley & Sons.

ISO 15007-1:2014, Road vehicles – Measurement of driver visual behaviour with respect to transport information and control systems – Part 1: Definitions and parameters. (2014). Geneva: International Organization for Standardization. Retrieved from `http://www.iso.org/iso/catalogue_detail.htm?csnumber=56621`

ISO 16673:2007, Road vehicles – Ergonomic aspects of transport information and control systems

– Occlusion method to assess visual demand due to the use of in-vehicle systems. (2007). Geneva: International Organization for Standardization. Retrieved from `http://www.iso.org/iso/catalogue_detail.htm?csnumber=38035`

ISO 17287:2003, Road vehicles – Ergonomic aspects of transport information and control systems – Procedure for assessing suitability for use while driving. (2003). Geneva: International Organization for Standardization. Retrieved from `http://www.iso.org/iso/home/store/catalogue_tc/catalogue_detail.htm?csnumber=30597`

ISO 17488:2016, Road vehicles – Transport information and control systems – Detection-Response Task (DRT) for assessing attentional effects of cognitive load in driving. (2016). Geneva: International Organization for Standardization. Retrieved from `http://www.iso.org/iso/home/store/catalogue_/catalogue_detail.htm?csnumber=59887`

ISO 26022:2010, Road vehicles – Ergonomic aspects of transport information and control systems – Simulated lane change test to assess in-vehicle secondary task demand. (2007). Geneva: International Organization for Standardization. Retrieved from `http://www.iso.org/iso/home/store/catalogue_tc/catalogue_detail.htm?csnumber=43361`

ISO/TS 14198:2012, Road vehicles – Ergonomic aspects of transport information and control systems – Calibration tasks for methods which assess driver demand due to the use of in-vehicle systems. (2012). Geneva: International Organization for Standardization. Retrieved from `http://www.iso.org/iso/home/store/catalogue_tc/catalogue_detail.htm?csnumber=54496`

ISO/TS 15007-2:2014, Road vehicles – Measurement of driver visual behaviour with respect to transport information and control systems – Part 2: Equipment and procedures. (2014). Geneva: International Organization for Standardization. Retrieved from `http://www.iso.org/iso/catalogue_detail.htm?csnumber=56622`

Jahn, G., Oehme, A., Krems, J. F., & Gelau, C. (2005). Peripheral detection as a workload measure in driving: Effects of traffic complexity and route guidance system use in a driving study. *Transportation Research Part F: Traffic Psychology and Behaviour*, *8*(3), 255–275. doi: 10.1016/j.trf.2005.04.009

James, W. (1890). *The principles of psychology* (electronic ed.). Retrieved from `http://psychclassics.yorku.ca/James/Principles/prin11.htm`

Jastrow, J. (1890). The time-relations of mental phenomena. *Science*, *16*(394), 99–101.

Jensen, O., & Mazaheri, A. (2010). Shaping functional architecture by oscillatory alpha activity: Gating by inhibition. *Frontiers in Human Neuroscience*, *4*. doi: 10.3389/fnhum.2010.00186

Johnston, W. A., & Dark, V. J. (1986). Selective Attention. *Annual Review of Psychology*, *37*, 43–75. doi: 10.1146/annurev.ps.37.020186.000355

Jones, L. A., & Sarter, N. B. (2008). Tactile displays: Guidance for their design and application. *Human Factors*, *50*(1), 90–111. doi: 10.1518/001872008X250638

Jonides, J. (1981). Voluntary versus automatic control over the mind's eye's movement. In J. Long & A. Baddeley (Eds.), *Attention and performance ix* (pp. 187–203). Hillsdale: Erlbaum. Retrieved from `http://wexler.free.fr/library.php`

Jonides, J., Schumacher, E. H., Smith, E. E., Lauber, E. J., Awh, E., Minoshima, S., & Koeppe, R. A. (1997). Verbal working memory load affects regional brain activation as measured by PET. *Journal of Cognitive Neuroscience*, *9*(4), 462–475. doi: 10.1162/jocn.1997.9.4.462

Klauer, S. G., Dingus, T. A., Neale, V. L., Sudweeks, J. D., & Ramsey, D. J. (2006). The impact of driver inattention on near crash/crash risk: An analysis using the 100-car naturalistic driving study data. (DOT HS 810 594), National Highway Traffic Safety Administration. Retrieved from `https://vtechworks.lib.vt.edu/bitstream/handle/10919/55090/DriverInattention.pdf?sequence=1`

Klimesch, W. (1999). EEG alpha and theta oscillations reflect cognitive and memory performance: a review and analysis. *Brain Research Reviews*, *29*(2-3), 169–195. doi: 10.1016/S0165-0173(98)00056-3

Klimesch, W., Sauseng, P., & Hanslmayr, S. (2007). EEG alpha oscillations: The inhibition–timing hypothesis. *Brain Research Reviews*, *53*(1), 63–88. Retrieved from `http://linkinghub.elsevier.com/retrieve/pii/S016501730600083X` doi: 10.1016/j.brainresrev.2006.06.003

Knowles, W. B. (1963). Operator loading tasks. *Human Factors*, *5*, 155–161. doi: 10.1177/001872086300500206

Kok, A. (2001). On the utility of P3 amplitude as a measure of processing capacity. *Psychophysiology*, *38*, 557–577. Retrieved from `http://onlinelibrary.wiley.com/wol1/doi/10.1017/S0048577201990559/abstract`

Krems, J. F., Keinath, A., Baumann, M., Gelau, C., & Bengler, K. (2000). Evaluating visual display designs in vehicles: Advantages and disadvantages of the occlusion technique. In L. M. Camarinha-Matos, H. Afsarmanesh, & H.-H. Erbe (Eds.), *Advances in networked enterprises: Virtual organizations, balanced automation, and systems integration* (pp. 361–368). Boston, Massachusetts: Springer. doi: 10.1007/978-0-387-35529-0_34

LaBerge, D. (1983). Spatial extent of attention to letters and words. *Journal of Experimental Psychology: Human Perception and Performance*, *9*(3), 371–379. doi: 10.1037/0096-1523.9.3.371

Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: A practical primer for t-tests and ANOVAs. *Frontiers in Psychology*, *4*, 1–12. doi: 10.3389/fpsyg.2013.00863

Lee, J. D., Young, K. L., & Regan, M. A. (2009). Defining driver distraction. In M. A. Regan, J. D. Lee, & K. L. Young (Eds.), *Driver distraction: Theory, effects, and mitigation* (pp. 31–40). Boca Raton, Florida: CRC.

Lee, J. Y., Gibson, M. C., & Lee, J. D. (2016). Error recovery in multitasking while driving. In *Proceedings of the 2016 chi conference on human factors in computing system* (pp. 7–12). New York, New York: ACM. Retrieved from `https://www.researchgate.net/profile/Ja_Young_Lee/publication/295854663_Error_Recovery_in_Multitasking_While_Driving/links/56ce0ede08ae059e37535549.pdf?inViewer=0{&}pdfJsDownload=0&origin=publication_detail`

Lei, S. (2011). *Driver mental states monitoring based on brain signals* (Doctoral thesis). Technical University of Berlin, Berlin, Germany.

Lei, S., & Roetting, M. (2011). Influence of task combination on EEG spectrum modulation for driver workload estimation. *Human Factors*, *53*(2), 168–179. doi: 10.1177/0018720811400601

Lin, C.-t., Chen, S.-a., Ko, L.-w., & Wang, Y.-k. (2011). EEG-based brain dynamics of driving distraction. In *Proceedings of international joint conference on neural networks* (pp. 1497–1500).

Liu, X. (2014). *Statistical power analysis for the social behavioral sciences: Basic and advanced techniques*. New York: Routledge.

Luce, R. (1986). *Response times: Their role in inferring elementary mental organization*. New York: Oxford University Press.

Luck, S. J., & Vecera, S. P. (2002). Attention. In H. Pashler & S. Yantis (Eds.), *Stevens' handbook of experimental psychology* (3rd ed., Vol. 1, pp. 235–286). New York: John Wiley & Sons.

Luck, S. J., Woodman, F., & Vogel, E. K. (2000). Event-related potential studies of attention. *Trends in Cognitive Science*, *4*(11), 432–440.

Mantzke, O., & Keinath, A. (2015). Relating the detection response task to critical events – consequences of high cognitive workload to brake reaction times. *Procedia Manufacturing*, *3*, 2381–2386. doi: 10.1016/j.promfg.2015.07.386

Marieb, E. N., & Hoehn, K. (2016). *Human Anatomy & Physiology* (10th ed.). Essex: Pearson Education Limited.

Mather, G. (2011). *Essentials of sensation and perception*. East Sussex: Routledge.

Mehler, B., Reimer, B., & Coughlin, J. F. (2012). Sensitivity of physiological measures for detecting systematic variations in cognitive demand from a working memory task: An on-road study across three age groups. *Human Factors*, *54*(3), 396–412. doi: 10.1177/0018720812442086

Mehler, B., Reimer, B., Coughlin, J. F., & Dusek, J. A. (2009). Impact of incremental increases in cognitive workload on physiological arousal and performance in young adult drivers. *Transportation Research Record: Journal of the Transportation Research Board*, *2138*, 6–12. doi: 10.3141/2138-02

Mehler, B., Reimer, B., & Zec, M. (2012). Defining workload in the context of driver state detection and HMI evaluation. In *Proceedings of the 4th international conference on automotive user interfaces and interactive vehicular applications - automotiveui '12* (pp. 187–192). doi: 10.1145/2390256.2390288

Merat, N., & Jamson, A. H. (2008). The effect of stimulus modality on signal detection: Implications for assessing the safety of in-vehicle technology. *Human Factors*, *50*(1), 145–158. doi: 10.1518/001872008X250656

Miller, E. K., & Cohen, J. D. (2001). An integrative theory of prefrontal cortex funtion. *Annual Review of Neuroscience*, *24*(1), 167–202. doi: 10.1146/annurev.neuro.24.1.167

Miura, T. (1986). Coping with situational demands: A study of eye movements and peripheral vision performance. In A. Gale, M. Freeman, C. Gaslegrave, P. Smith, & S. Taylor (Eds.), *Vision in vehichles: Proceedings of the conference on vision vehicles* (pp. 205–216). Amsterdam: Elsevier.

Mole, C. (2012). Attention. In E. Margolis, R. Samuels, & S. Stich (Eds.), *The oxford handbook of philosophy of cognitive science* (pp. 201–221). New York: Oxford University Press.

Müller, M. M., & Giabbiconi, C.-M. (2008). Attention in sense of touch. In G. Martin (Ed.), *Human haptic perception: Basics and applications* (pp. 199–206). Basel: Birkhäuser.

Navon, D., & Gopher, D. (1979). On the economy of the human-processing system. *Psychological Review*, *86*(3), 214–255. doi: 10.1037/0033-295X.86.3.214

Ng, A. W. Y., & Chan, A. H. S. (2012). Finger response times to visual, auditory and tactile modality stimuli. In *Proceedings of the international multiconference of engineers and computer scientists* (Vol. II, pp. 1449–1454). Retrieved from `http://www.iaeng.org/publication/IMECS2012/IMECS2012_pp1449-1454.pdf`

Ninio, A., & Kahneman, D. (1974). Reaction time in focused and in divided attention. *Journal of Experimental Psychology*, *103*(3), 394–399. doi: 10.1037/h0037202

Norman, D. A., & Bobrow, D. G. (1975). On data-limited and resource-limited processes. *Cognitive Psychology*, *7*(1), 44–64. doi: 10.1016/0010-0285(75)90004-3

Norman, D. A., & Bobrow, D. G. (1976). On the analysis of performance operating characteristics. *Psychological Review*, *83*(6), 508–510. doi: 10.1037/0033-295X.83.6.508

O'Donell, R., & Eggemeier, T. (1986). Workload assessment methodology. In K. R. Boff, L. Kaufman, & J. P. Thomas (Eds.), *Handbook of perception and human performance* (pp. 42–1 – 42–49). New York: Wiley.

Ogden, G. D., Levine, J. M., & Eisner, E. J. (1979). Measurement of workload by secondary tasks. *Human Factors*, *21*(5), 529–548.

Olsson, S., & Burns, P. C. (2000). Measuring driver visual distraction with a peripheral detection task. In *Nhtsa driver distraction internet forum.* Retrieved from `http://www.nhtsa.gov`

Östlund, J., Nilsson, L., Carsten, O., Merat, N., Jamson, H., Jamson, S., … Brouwer, R. (2004). HASTE deliverable 2: HMI and safety-related driver performance.

Parasuraman, R., & Caggiano, D. (2002). Mental workload. In V. Ramachandran (Ed.), *Encyclopedia of the human brain* (Vol. 3, pp. 17–27). San Diego: Academic Press.

Pashler, H. E. (1998a). Capacity and selection: Theorizing about attention. In *The psychology of attention* (pp. 217–261). Cambridge: MIT Press.

Pashler, H. E. (1998b). *The psychology of attention.* Cambridge: MIT Press.

Patten, C. J. D., Kircher, A., Östlund, J., Nilsson, L., & Svenson, O. (2006). Driver experience and cognitive workload in different traffic environments. *Accident Analysis and Prevention*, *38*(5), 887–894. doi: 10.1016/j.aap.2006.02.014

Pauzié, A. (2008). A method to assess the driver mental workload: The driving activity load index (DALI). *IET Intelligent Transport Systems*, *2*(4), 315. doi: 10.1049/iet-its:20080023

Petzoldt, T., Weiß, T., Franke, T., Krems, J. F., & Bannert, M. (2013). Can driver education be improved by computer based training of cognitive skills? *Accident Analysis and Prevention*, *50*, 1185–1192. doi: 10.1016/j.aap.2012.09.016

Posner, M. I., & Boies, S. J. (1971). Components of attention. *Psychological Review*, *78*(5), 391–408. doi: 10.1037/h0031333

Proctor, R. W., & Proctor, J. D. (2006). Sensation and Perception. In G. Salvendy (Ed.), *Handbook of human factors and ergonomics* (3rd ed., pp. 53–88). Hoboken, New Jersey: John Wiley

& Sons.

Proctor, R. W., & Vu, K.-P. L. (2003). Action selection. In I. B. Weiner, A. F. Healy, & R. W. Proctor (Eds.), *Handbook of psychology* (Vol. 4, pp. 293–316). Hoboken, New Jersey: John Wiley & Sons.

Purves, D., Augustine, G., Fitzpatrick, D., Hall, W. C., LaMantia, A.-S., & White, L. E. (2001). *The somatic sensory cortex.* Retrieved 2016-02-23, from `http://www.ncbi.nlm.nih.gov/books/NBK11153/`

Ranney, T., Garrott, W., & Goodman, M. (2000). NHTSA Driver Distraction Research: Past, Present, and Future. Retrieved from `http://www-nrd.nhtsa.dot.gov/pdf/esv/esv17/proceed/00177.pdf`

Ranney, T. A. (2008). Driver Distraction: A Review of the Current State-of-Knowledge. (Report No. DOT HS 810 787), National Highway Traffic Safety Administration.

Ranney, T. A., Baldwin, G., Smith, L. A., Mazzae, E., & Pierce, R. S. (2014). Detection response task evaluation for driver distraction measurement application. (Report No. DOT HS 812 077), National Highway Traffic Safety Administration.

Richardson, J. T. E. (2011). Eta squared and partial eta squared as measures of effect size in educational research. *Educational Research Review*, *6*(2), 135–147. doi: 10.1016/j.edurev.2010.12.001

Robinson, D. K. (2001). Reaction-time experiments in Wundt's institute and beyond. In R. W. Rieber & D. K. Robinson (Eds.), *Wilhelm wundt in history: The making of a scientific psychology* (pp. 161–204). New York: Kluwer Academic / Plenum Publishers.

Rouse, W., Edwards, S., & Hammer, J. (1993). Modeling the dynamics of mental workload and human performance in complex systems. *IEEE Transactions on Systems, Man, and Cybernetics*, *23*(6), 1662–1671. doi: 10.1109/21.257761

Rumelhart, D., Hinton, G., & McClelland, J. (1986). A General framework for Parallel Distributed Processing. In D. Rumelhart & J. McClelland (Eds.), *Parallel distributed processing: explorations in the microstructure of cognition* (Vol. 1, pp. 45 – 76). Cambridge: MIT Press.

SAE J 2944 - Operational definitions of driving performance measures and statistics. (2015). SAE International. Retrieved from `http://standards.sae.org/j2944_201506/`

Sammer, G. (2001). Workload and electro-encephalography, information processing and workload. In W. Karwowski (Ed.), *International encyclopedia of ergonomics and human factors* (pp. 350–352). London: Taylor & Francis.

Sauseng, P., Hoppe, J., Klimesch, W., Gerloff, C., & Hummel, F. C. (2007). Dissociation of sustained attention from central executive functions: Local activity and interregional connectivity in the theta range. *European Journal of Neuroscience*, *25*(2), 587–593. doi: 10.1111/j.1460-9568.2006.05286.x

Sauseng, P., & Klimesch, W. (2008). What does phase information of oscillatory brain activity tell us about cognitive processes? *Neuroscience and Biobehavioral Reviews*, *32*(5), 1001–1013. doi: 10.1016/j.neubiorev.2008.03.014

Schaap, T. W., van der Horst, A. R. A., van Arem, B., & Brookhuis, K. (2009). The relationship between driver distraction and mental workload. In *First international conference on driver distraction and inattention (ddi 2009).* Retrieved from `https://trid.trb.org/view.aspx?id=1240040`

Schmider, E., Ziegler, M., Danay, E., Beyer, L., & Bühner, M. (2010). Is it really robust? Reinvestigating the robustness of ANOVA against violations of the normal distribution assumption. *Methodology*, *6*(4), 147–151. doi: 10.1027/1614-2241/a000016

Shinar, D. (2015). Cognitive workload ≠ crash risk: Rejoinder to study by Strayer et al. (2015). *Human Factors*, *57*(8), 1328–1330. doi: 10.1177/0018720815574943

Shinar, D., & Vogelzang, M. (2013). Comprehension of traffic signs with symbolic versus text displays. *Transportation Research Part F: Traffic Psychology and Behaviour*, *18*, 72–82. doi: 10.1016/j.trf.2012.12.012

Siam, M. F. M., Isa, M. H. M., Sukardi, A., Borhan, N., & Voon, W. S. (2014). *Measuring attention and performance in simulated road traffic environments using detection response task method* (No. MRR No.150). Selangor: Malaysian Institute of Road Safety Research (MIROS).

Speckmann, E., & Elger, C. (2005). Introduction to the neurophysiological basis of the EEG and DC potentials. In E. Niedermeyer & F. Lopes da Silva (Eds.), *Electroencephalography: Basic principles, clinical applications, and related fields* (5th ed., pp. 17–29). Philadelphia: Lippincott Williams & Wilkins.

Steriade, M. (2005). Cellular substrates of brain rhythms. In E. Niedermeyer & F. Lopes da Silva (Eds.), *Electroencephalography: Basic principles, clinical applications, and related fields* (5th ed., pp. 31–83). Philadelphia: Lippincott Williams & Wilkins.

Strayer, D. L., Cooper, J., Turrill, J., Coleman, J. R., & Hopman, R. J. (2015). *Measuring cognitive distraction in the automobile III: A comparison of Ten 2015 in-vehicle information systems*. Washington, DC: AAA Foundation for Traffic Safety.

Strayer, D. L., Cooper, J. M., & Drews, F. A. (2004). What do drivers fail to see when conversing on a cell phone? In *Proceedings of the human factors and ergonomics society 48th annual meeting* (pp. 2213–2217). doi: 10.1177/154193120404801902

Strayer, D. L., & Drews, F. A. (2007). Cell-phone–induced driver distraction. *Current Directions in Psychological Science*, *16*(3), 128–131.

Strayer, D. L., & Fisher, D. L. (2016). SPIDER: A framework for understanding driver distraction. *Human Factors*, *58*(1), 5–12. doi: 10.1177/0018720815619074

Strayer, D. L., & Johnston, W. A. (2001). Driven to distraction: Dual-task studies of simulated driving and conversing on a cellular telephone. *Psychological Science*, *12*(6), 462–466. doi: 10.1111/1467-9280.00386

Strayer, D. L., Turrill, J., Coleman, J. R., Ortiz, E. V., & Cooper, J. M. (2014). *Measuring cognitive distraction in the automobile II: assessing in-vehicle voice-based interactive technologies*. Washington, DC: AAA Foundation for Traffic Safety.

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, *18*(6), 643–662. doi: 10.1037/h0054651

Sullivan, G. M., & Feinn, R. (2012). Using effect size - or why the P value is not enough. *Journal of Graduate Medical Education*, *4*(3), 279–82. doi: 10.4300/JGME-D-12-00156.1

Summala, H. (1996). Accident risk and driver behaviour. *Safety Science*, *22*(1-3), 103–117. doi: 10.1016/0925-7535(96)00009-4

Summala, H. (2007). Towards understanding motivational and emotional factors in driver behavior: Comfort through satisficing. In P. C. Cacciabue (Ed.), *Modelling driver behavior in automotive environments: critical issues in driver interactions with intelligent transport systems* (pp. 189–207). London: Springer.

Teplan, M. (2002). Fundamentals of EEG measurement. *Measurement Science Review*, *2*, 1–11.

Theeuwes, J. (1992). Perceptual selectivity for color and form. *Perception & Psychophysics*, *51*(6), 599–606. doi: 10.3758/BF03211656

Theeuwes, J., Kramer, A. F., Hahn, S., & Irwin, D. E. (1998). Our eyes do not always go where we want them to go: Capture of the eyes by new objects. *Psychological Science*, *9*(5), 379–385. doi: 10.1111/1467-9280.00071

Tijerina, L. (2000). Issues in the evaluation of driver distraction associated with in-vehicle information and telecommunications systems.. Retrieved from `http://www-nrd.nhtsa.dot.gov/departments/Human%20Factors/driver-distraction/PDF/3.PDF`

Treisman, A. (1985). Preattentive processing in vision. *Computer Vision, Graphics, and Image Processing*, *31*(2), 156–177. doi: 10.1016/S0734-189X(85)80004-9

Treisman, A. M. (1960). Contextual cues in selective listening. *Quarterly Journal of Experimental Psychology*, *12*(4), 242–248. doi: 10.1080/17470216008416732

Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology*, *12*(1), 97–136. doi: 10.1016/0010-0285(80)90005-5

Trick, L. M., & Enns, J. T. (2009). Framework for understanding the role of attentional selection in driving. In C. Castro (Ed.), *Human factors of visual and cognitive performance in driving* (pp. 63–73). Boca Raton, Florida: CRC Press.

Tsang, P. S., & Vidulich, M. A. (2006). Mental workload and situation awareness. In G. Salvendy (Ed.), *Handbook of human factors and ergonomics* (Third ed., pp. 243–268). Hoboken, New Jersey: John Wiley & Sons.

Ünal, A. B., Platteel, S., Steg, L., & Epstude, K. (2013). Blocking-out auditory distracters while driving: A cognitive strategy to reduce task-demands on the road. *Accident Analysis and Prevention*, *50*, 934–942. doi: 10.1016/j.aap.2012.07.020

van der Horst, A. R. A., & Martens, M. H. (2010). The peripheral detection task (PDT): On-line measurement of driver cognitive workload and selective attention. In G. L. Rupp (Ed.), *Performance metrics for assessing driver distraction: The quest for improved road safety* (pp. 73 – 89). Warrendale: SAE International.

van Winsum, W., Martens, M., & Herland, L. (1999). *The effects of speech versus tactile driver support messages on workload, driver behaviour and user acceptance.* Soesterberg: TNO Human Factors Research Institute (Report No. TM-99-C043).

Vollrath, M., Huemer, A. K., Teller, C., Likhacheva, A., & Fricke, J. (2016). Do German drivers use their smartphones safely?—Not really! *Accident Analysis & Prevention*, *96*, 29–38. doi: 10.1016/j.aap.2016.06.003

Wässle, H. (2004). Parallel processing in the mammalian retina. *Nature Reviews Neuroscience*, *5*(10), 747–757. doi: 10.1038/nrn1497

Weisz, N., Wühle, A., Monittola, G., Demarchi, G., Frey, J., Popov, T., & Braun, C. (2014). Prestimulus oscillatory power and connectivity patterns predispose conscious somatosensory perception. *Proceedings of the National Academy of Sciences*, *111*(4), E417–E425. doi: 10.1073/pnas.1317267111

Wickens, C. D. (2002). Multiple resources and performance prediction. *Theoretical Issues in Ergonomics Science*, *3*(2), 159–177. doi: 10.1080/14639220210123806

Wickens, C. D., & Hollands, J. (2000). Attention, time-sharing, and workload. In *Engineering psychology and human performance* (3rd ed., pp. 439–479). Upper Saddle River, NJ: Prentice Hall.

Wickens, C. D., Hollands, J. D., Banbury, S., & Parasuraman, R. (2013a). *Engineering psychology and human performance* (International; 4th ed.). Boston: Pearson.

Wickens, C. D., Hollands, J. G., Banbury, S., & Parasuraman, R. (2013b). Multitasking corrected. In *Engineering psychology and human performance* (International; 4th ed., chap. 10). Boston: Pearson.

Wickens, C. D., Isreal, J., & Donchin, E. (1977). The event related cortical potential as an index of task workload. In *Proceedings of the human factors and ergonomics society 21st annual meeting* (pp. 282–286).

Wickens, C. D., & McCarley, J. S. (2008). *Applied attention theory.* Boca Raton, Florida: CRC.

Young, K., Regan, M., & Hammer, M. (2003). *Driver distraction: A review of the literature.* Monash University Accident Research Centre. Retrieved from `http://www.monash.edu.au/miri/research/reports/muarc206.pdf`

Young, K., Regan, M., & Hammer, M. (2007). Driver distraction: A review of the literature. In I. Faulks, M. Regan, M. Stevenson, J. Brown, A. Porter, & J. Irwin (Eds.), *Distracted driving* (pp. 379–405). Sydney: Australasian College of Road Safety.

Young, M. S., Brookhuis, K. A., Wickens, C. D., & Hancock, P. A. (2015). State of science: Mental workload in ergonomics. *Ergonomics*, *58*(1), 1–17. doi: 10.1080/00140139.2014.956151

Young, R. A. (2015). Driver compensation: Impairment or improvement? *Human Factors*, *57*(8), 1334–1338. doi: 10.1177/0018720815585053

Young, R. A., Hsieh, L., & Seaman, S. (2013). The tactile detection response task: Preliminary validation for measuring the attentional effects of cognitive load. In *Proceedings of the seventh international driving symposium on human factors in driver assessment, training, and vehicle design* (pp. 71–77). Iowa City: University of Iowa. Retrieved from `http://drivingassessment.uiowa.edu/sites/default/files/DA2013/Papers/012_Young_0.pdf`

Appendix A

Supplementary material for Experiment I

A.1. Sentences task script

Sentences task for Experiment I per difficulty level (easy, hard) and topic (geography, sports, biology, physical, relations, food, personal).

Easy level

- Geography
  - Rom ist eine Stadt in Italien.
  - München ist die Hauptstadt von Österreich.
  - Hamburg liegt im Norden Deutschlands.
  - Japan ist weiter weg als Spanien.
  - Die Zugspitze ist der höchste Berg in Deutschland.
  - Deutschland ist eine Insel.
  - England liegt in Südeuropa.
  - Afrika ist kleiner als Holland.

  - Die Hauptstadt von Amerika ist Mexiko.
  - Nürnberg liegt im Norden Bayerns.
  - Kanada liegt in Nordamerika.
  - Mallorca ist eine spanische Insel.

- Sport
  - Für Tennis benötigt man einen Schläger.
  - Beim Fußball spielen zwei Mannschaften gegeneinander.
  - Beim Schwimmen wird man nass.
  - Ein zwei Euro Stück ist kleiner als ein Cent Stück.
  - Meerwasser ist süß.
  - Ein Würfel hat 15 Ecken.
  - Die Erde ist eine Scheibe.
  - Schwere Dinge fallen schneller als leichte.
  - Ein normales Fahrrad hat 5 Räder.
  - Eishockey wird auf dem Eis gespielt.
  - Fechten ist eine olympische Disziplin.
  - Lufthansa ist eine deutsche Fluggesellschaft.

- Biology
  - Ein Huhn hat zwei Beine.
  - Ein Pinguin kann nicht fliegen.
  - Ein Hund hat sechs Beine.
  - Im Herbst werden die Blätter grün.
  - Schlangen haben vier Beine.

    - Pflanzen benötigen Wasser zum Wachsen.
    - Katzen sind Raub tiere.
    - Fledermäuse fliegen normalerweise nachts.
    - Eis Bären halten einen Winterschlaf.
    - Mäuse fressen Katzen.
    - Elefanten legen Eier.
    - Kamele leben in der Arktis.
- Physical
    - Wasser gefriert bei etwa 100 Grad Celsius.
    - Wasser kocht bei etwa 0 Grad Celsius.
    - Schwarz ist dunkler als weiß.
    - Kork schwimmt auf Wasser.
    - Helium ist ein sehr leichtes Gas.
    - Kalte Luft steigt nach oben.
    - Es schneit im Sommer.
    - Ein Wasser molekül besteht aus 3 Atomen.
    - Gravitation ist eine Kraft.
    - Eisen ist weicher als Butter.
    - Kupfer ist ein Isolator.
    - Kraft ist das Produkt aus Masse mal Beschleunigung.
    - Ohne Gravitation würde man schweben.
- Relations
    - Die Mutter meiner Mutter ist meine Großmutter.

  - Der Sohn meiner Schwester ist mein Neffe.
  - Der Bruder meines Vaters ist mein Onkel.
  - Die Kinder meiner Kinder sind meine Schwestern.
  - Die Kinder meiner Tante sind meine Kinder.
  - Der Mann meiner Schwester ist mein Bruder.
  - Die Mutter meines Bruder ist meine Mutter.
  - Der Vater meiner Cousine ist mein Onkel.
  - Die Kinder meiner Schwester sind meine Nichten und Neffen.
  - Der Sohn meines Sohns ist meine Tochter.
  - Die Tochter meiner Tante ist mein Vetter.
  - Der Bruder meiner Schwester ist mein Vater.

- ○ Food
  - Nudeln sind langes dünnes Gemüse.
  - Bratwurst ist eine chinesisch Spezialität.
  - Suschi wird mit rohem Fisch zubereitet.
  - Reis kann in Wasser gekocht werden.
  - Pommes Frits werden aus Fleisch gemacht.
  - Kaffee ist ein Heißgetränk.
  - Äpfel wachsen auf Bäumen.
  - Bananen sind eine gelbe tropische Frucht.
  - Milch kommt meist von Kühen.
  - Wiener Wald ist ein Delikatessen Restaurant.
  - Eis ist ein grünes Gemüse.

  - Spinat wird normaler weise als Dessert serviert.

- Personal
  - Ihr Name ist Angela Merkel.
  - Sie sind in Australien geboren.
  - Sie sind eine weibliche Versuchsperson.
  - Sie sind eine männliche Versuchsperson.
  - Sie sind unter 40 Jahre alt.
  - Sie haben Ihren Führerschein gemacht.
  - Sie befinden sich aktuell im Maschinenbau Gebäude.
  - Sie sind derzeit mitten in einem Experiment.
  - Das heutige Wetter ist schlecht.
  - Heute regnet es.
  - Heute scheint die Sonne.
  - Sie sind ein Profi Fußball Spieler.

Hard level

- Geography
  - Rom ist eine Italien Stadt in.
  - München ist die Hauptstadt Österreich von.
  - Hamburg Norden liegt Deutschlands im.
  - Japan ist weiter Spanien weg als.
  - Die Zugspitze ist der höchste Deutschland in Berg.
  - Deutschland ist Insel eine.
  - England Südeuropa liegt in.

  - Afrika ist als kleiner Holland.
  - Die Hauptstadt von Amerika Mexiko ist.
  - Liegt Nürnberg im Norden Bayerns.
  - Kanada liegt Nordamerika in.
  - Mallorca ist eine Insel spanische.

- Sport
  - Für Tennis benötigt Schläger man einen.
  - Beim Fußball spielen Mannschaften 2 gegeneinander.
  - Beim Schwimmen nass man wird.
  - Ein 2 € Stück ist kleiner Stück als ein Cent.
  - Meerwasser süß ist.
  - Ein Würfel Ecken 15 hat.
  - Die Erde eine Scheibe ist.
  - Schwere Dinge schneller fallen als leichte.
  - Ein normales Fahrrad hat Räder fünf.
  - Eishocky auf Eis wird gespielt.
  - Fechten eine olympische Disziplin ist.
  - Lufthansa ist deutsche eine Fluggesellschaft.

- Biology
  - Ein Huhn zwei hat Beine.
  - Ein Pinguin fliegenkann nicht.
  - Ein Hund sechs hat Beine.
  - Im Herbst Blätter werden die grün.

  - Schlangen vier haben Beine.
  - Pflanzen benötigen Wasser Wachsen zum.
  - Katzen Raubtiere sind.
  - Fledermäuse normalerweise nachts fliegen.
  - Eis Bären halten Winterschlaf einen.
  - Mäuse Katzen fressen.
  - Elefanten Eier legen.
  - Kamele in der leben Arktis.

- ○ Physical
  - Wasser 100 Grad Celsius gefriert etwa bei.
  - Wasser kocht 0 Grad etwa bei.
  - Schwarz ist dunkler weiß als.
  - Kork schwimmt Wasser auf.
  - Helium ist ein leichtes sehr Gas.
  - Kalte Luft nach steigt oben.
  - Ein Wasser molekül aus besteht 3 Atomen.
  - Gravitation eine ist Kraft.
  - Eisen weicher ist als Butter.
  - Kupfer ist Isolator ein.
  - Kraft ist das Produkt Beschleunigung aus Masse mal.
  - Ohne Gravitation schweben würde man.
- ○ Relations
  - Die Mutter meiner Mutter ist Großmutter meine.

  - Der Sohn meiner Schwester mein Neffe ist.
  - Der Bruder meines Vaters ist Onkel mein.
  - Die Kinder meiner Kinder meine sind Schwester.
  - Die Kinder meiner Tante meine Kinder sind.
  - Der Mann meiner Schwester mein ist Bruder.
  - Die Mutter meines Bruder ist Mutter meine.
  - Der mein Vater meiner Cousine ist Onkel.
  - Die Kinder meiner Schwester sind Nichten und Neffen meine.
  - Der Sohn meines Sohn ist Tochter meine.
  - Die Tochter meiner Tante mein ist Vater.
  - Der Bruder meiner Schwester mein ist Vater.

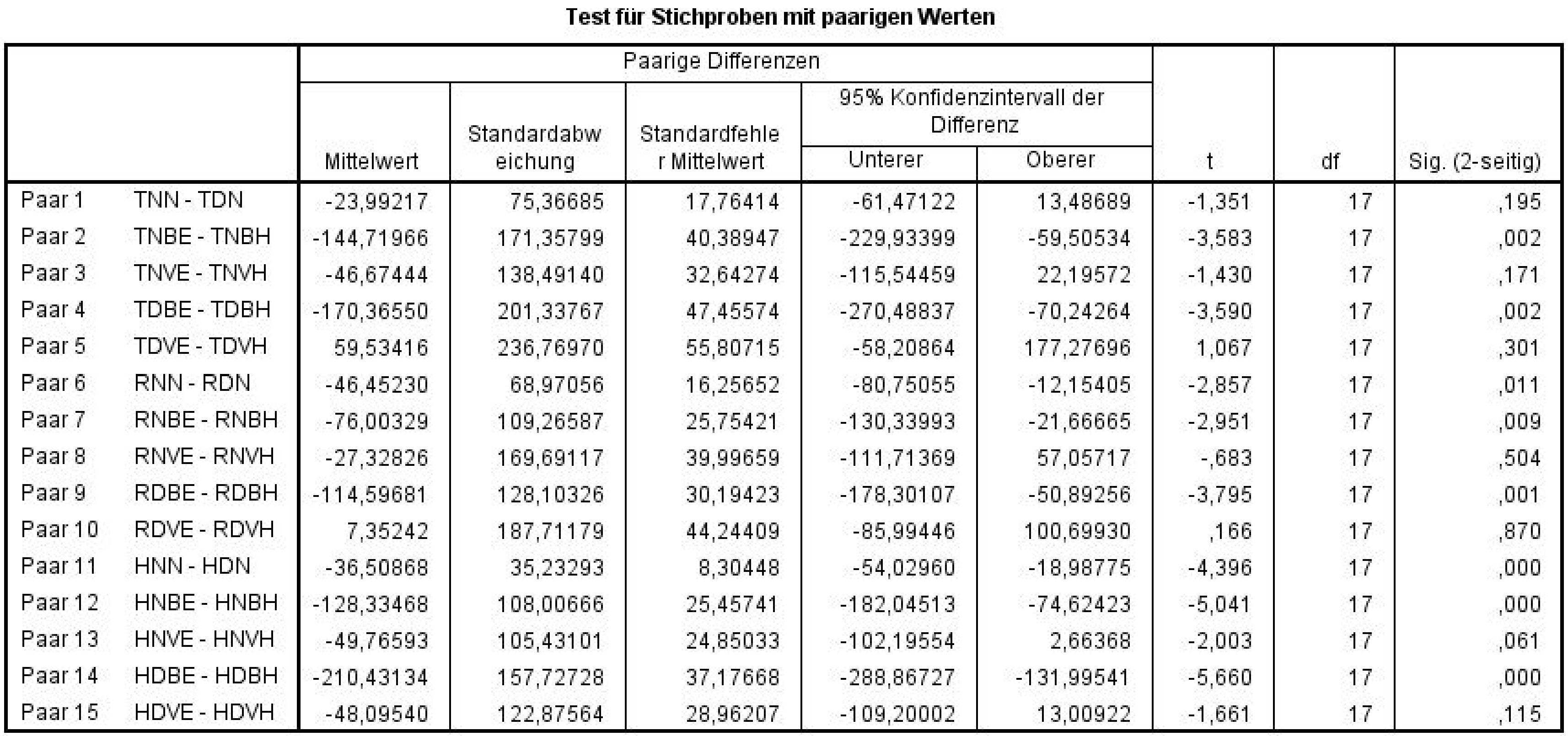
- ○ Food
  - Nudeln sind langes Gemüse dünnes.
  - Bratwurst ist Chinesisch eine Spezialität.
  - Suschi wird mit rohem zubereitet Fisch.
  - Reis kann in gekocht Wasser werden.
  - Pommes Frits werden aus gemacht Fleisch.
  - Kaffee Heißgetränke ist ein.
  - Äpfel wachsen Bäumen auf.
  - Bananen sind gelbe eine tropische Frucht.
  - Milch kommt meist Kühen von.
  - Wiener Wald ist Delikatessen ein Restaurant.
  - Eis grünes Gemüse ist ein.

  - Spinat wird als Dessert normaler weise serviert.

- Personal
  - Ihr ist Name Angela Merkel.
  - Sie sind in geboren Australien.
  - Sie sind eine Versuchsperson weibliche.
  - Sie sind männliche eine Versuchsperson.
  - Sie sind unter Jahre alt vierzig.
  - Sie haben Ihren gemacht Führerschein.
  - Sie befinden sich aktuell im Gebäude Maschinen bau.
  - Sie sind mitten in einem der zeit Experiment.
  - Das Wetter heutige schlecht ist.
  - Es heute regnet.
  - Heute Sonne scheint die.
  - Sie sind ein Spieler Fußball Profi.

A.2. Results: *t*-test

(next page)



Test für Stichproben mit paarigen Werten

		Paarige Differenzen					t	df	Sig. (2-seitig)
		Mittelwert	Standardabweichung	Standardfehler Mittelwert	95% Konfidenzintervall der Differenz				
					Unterer	Oberer			
Paar 1	TNN - TDN	-23,99217	75,36685	17,76414	-61,47122	13,48689	-1,351	17	,195
Paar 2	TNBE - TNBH	-144,71966	171,35799	40,38947	-229,93399	-59,50534	-3,583	17	,002
Paar 3	TNVE - TNVH	-46,67444	138,49140	32,64274	-115,54459	22,19572	-1,430	17	,171
Paar 4	TDBE - TDBH	-170,36550	201,33767	47,45574	-270,48837	-70,24264	-3,590	17	,002
Paar 5	TDVE - TDVH	59,53416	236,76970	55,80715	-58,20864	177,27696	1,067	17	,301
Paar 6	RNN - RDN	-46,45230	68,97056	16,25652	-80,75055	-12,15405	-2,857	17	,011
Paar 7	RNBE - RNBH	-76,00329	109,26587	25,75421	-130,33993	-21,66665	-2,951	17	,009
Paar 8	RNVE - RNVH	-27,32826	169,69117	39,99659	-111,71369	57,05717	-,683	17	,504
Paar 9	RDBE - RDBH	-114,59681	128,10326	30,19423	-178,30107	-50,89256	-3,795	17	,001
Paar 10	RDVE - RDVH	7,35242	187,71179	44,24409	-85,99446	100,69930	,166	17	,870
Paar 11	HNN - HDN	-36,50868	35,23293	8,30448	-54,02960	-18,98775	-4,396	17	,000
Paar 12	HNBE - HNBH	-128,33468	108,00666	25,45741	-182,04513	-74,62423	-5,041	17	,000
Paar 13	HNVE - HNVH	-49,76593	105,43101	24,85033	-102,19554	2,66368	-2,003	17	,061
Paar 14	HDBE - HDBH	-210,43134	157,72728	37,17668	-288,86727	-131,99541	-5,660	17	,000
Paar 15	HDVE - HDVH	-48,09540	122,87564	28,96207	-109,20002	13,00922	-1,661	17	,115

Figure A.1. – Results from *t*-test run for the power analysis of experiment I (N = 18), starting on page 73.

Appendix B

Supplementary material for Experiment II

B.1. Outliers

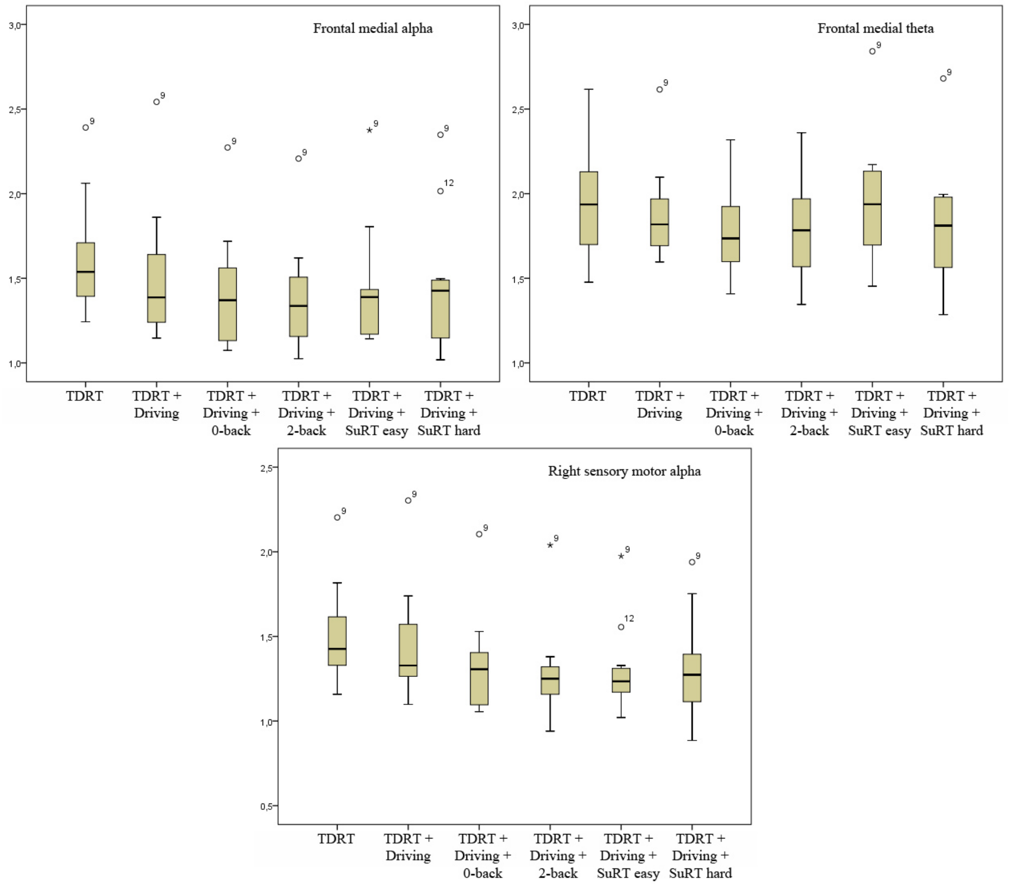


Figure B.1. – Outliers present in EEG data of Experiment II. Figures plot wave amplitudes (μV) as a function of condition.

B.2. Analysis of EEG data for secondary task baselines

Secondary task	EEG Metric	*M*	*SD*
n-back easy (0)	RSM Alpha	1.55	.34
	FM Alpha	1.58	.36
	FM Theta	1.79	.23
n-back hard (2)	RSM Alpha	1.50	.35
	FM Alpha	1.57	.41
	FM Theta	1.84	.32
SuRT easy	RSM Alpha	1.35	.22
	FM Alpha	1.47	.32
	FM Theta	2.12	.39
SuRT hard	RSM Alpha	1.44	.24
	FM Alpha	1.55	.34
	FM Theta	2.06	.29

Table B.1. – *M*s and *SD*s for EEG amplitudes for secondary task baselines (N = 12).

An analysis was implemented to establish whether EEG metrics were sensitive to load differences of n-back and SuRT difficulty levels. Three RM ANOVAs were used to determine whether statistically significant mean differences between easy and difficult secondary task variants were found in right sensory motor (RSM) alpha, frontal medial (FM) alpha, and FM theta. *M*s and *SD*s can be found in Table B.1. No outliers were excluded from the analysis. The normality assumption was not violated, as assessed by Shapiro-Wilk's test ($p > .05$). For RSM alpha, Mauchley's test of sphericity indicated that the assumption of sphericity had been violated, $\chi^2(5) = 17.28$, $p = .004$, therefore degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity ($\varepsilon = .505$). No significant effect of task load was found, $p = .059$. For FM alpha, no significant effect of task load was found, $p = .163$. For FM theta, Mauchley's test of sphericity indicated that the assumption of sphericity had been violated, $\chi^2(5) = 13.85$, $p = .017$, therefore degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity ($\varepsilon = .519$). Task load elicited statistically significant changes in mean amplitudes, $F(1.56, 17.13) = 11.98$, $p = .001$, partial $\eta^2 = .521$. Bonferroni corrected post-hoc analyses were performed to test for significant differences in mean RDRT RTs between each easy and difficult variant of secondary task, however, no relevant (easy task variant versus hard task variant for each secondary tasks) difficulty level comparisons were significant.