universität wien

Diplomarbeit

Titel der Diplomarbeit

Detection Response Tasks – A New Method to Measure Cognitive Workload

Verfasserin

Juliane Schäfer

Angestrebter akademischer Grad

Magistra der Naturwissenschaften (Mag. rer. nat.)

Wien, im Mai 2012

Studienkennzahl: 298

Studienrichtung: Psychologie

Betreuer: Univ.-Prof. Dipl.-Psych. Dr. Helmut Leder

Acknowledgement

First of all I would like to thank Prof. Dr. Helmut Leder for having made this diploma thesis possible, in cooperation between the University of Vienna and the BMW Group in Munich. Thank you.

Furthermore I would like to thank Dr. Roman Vilimek for his dedicated support, his valuable advice and for being generous in his offer of both advise and practical assistance when ever I have needed either one.
I also owe a big 'thank you' to my colleagues for the constructive discussions and the professional tips which contributed a great deal into making my time spent at the BMW such a pleasant experience.

A special thank goes to Faye for her expert view on my English.
And last but not least I want to thank my mum for her love and support.

Contents

I Preface

This thesis has been formulated in the English language for the specific purpose of publication in an English journal. For this reason, this article attempts to meet all prerequisites necessary and demanded of such scientific manuscripts, above everything, to remain precise, up to date and relevant to the theme it seeks to explore.

II Article

Detection Response Tasks – A New Method to Measure Cognitive Workload

Abstract

Currently, many new in-vehicle information and communication systems are entering our cars. They range from media players to navigation systems to speech based systems which allow the use of the vehicles functions by speech. While these systems are intended to facilitate the driving task, some research shows they also may also have the potential to distract drivers (e.g. Santos et al., 2005). In order to evaluate to which degree such systems are suitable for the while driving, many evaluation methods have been proposed. While most of these evaluation methods allow the measurement of visual-manual distraction (e.g. eye tracking), cognitive distraction is much more complicated to measure. However, as more systems become multi-modal or purely speech-based, it is getting even more important to measure the impact of cognitive distractions. Recent studies (Merat & Jamson, 2008) explored a promising method to evaluate cognitive workload: detection response tasks (DRTs). Such detection response tasks rely on at least a dual task setting, where the impairment in a secondary task (the detection response task) is an indication of the workload imposed by the primary task.

In this study three types of DRTs were evaluated: peripheral detection response task (PDRT), auditory detection response task (ADRT) and tactile detection response task (TDRT). In order to evaluate the sensitivity of each of these DRTs, cognitive tasks like the n-back task (Mehler et al., 2009) and a counting task were deployed in two levels of difficulty. Additionally, these cognitive tasks were presented in a visual, auditory, as well as pure cognitive way to clarify if any interactions between the different modalities of the DRTs and the presentation modes of the cognitive tasks exist.

Results revealed significant differences between high and low levels of cognitive workload for all three types of the DRT variants evaluating the reaction time. However, a closer examination of the results showed that the PDRT is not adequately sensitive to measure increased cognitive workload on the counting task if the dependent measure is the hit rate.

It is concluded that all three DRT variants are a sensitive measurement technique to assess cognitive workload. More research is needed to validate these findings on the use of real world tasks. Furthermore it has to be proven if it is possible to apply one of the

DRT variants to a tertiary design (driving + test task + DRT), as this would increase the ecological validity of the method.

Keywords: DRT; PDRT; ADRT; TDRT; N-back task; Cognitive workload; Driver distraction

1. Introduction

In recent years, the installation and use of various driver information systems, such as driver support and infotainment systems, has become more and more widespread in vehicles. These systems have been designed in order to deliver useful information to the driver and simplify the drive (navigation systems are a good example of this incentive). Nevertheless, the use of such systems has since been linked with the exercise of mental workload and the possibility of distraction. The degree to which those systems interfere with a drivers`attention and lead to distraction has yet to be determined. An even more precise estimate of these factors could be very beneficial, as it would have tremendous significance when it comes to ensuring a safer way of use of such systems while driving. Accident likelihood may increase if drivers engage in too many secondary tasks. Wierwille and Tijerina (1998) found evidence for the negative influence of the amount and frequency of visual attention to in-vehicle devices on the safety of drivers using a large pool of accident data.

In general, driver distraction can be categorized in the following three types:

- visual distraction: caused by tasks that require the driver to look away from the road to visually obtain information
- manual distraction: caused by tasks that require the driver to remove a hand from the steering wheel to manipulate a device
- cognitive distraction: caused by tasks that require the driver to avert their mental attention away from the driving task

Most of present day in-vehicle information systems are visual-manual systems as they use touch screens or controllers as input devices while they present information on display screens. As a consequence, most of the research has been concerned with visual-manual distraction (Angell et al., 2006). Recently this has also lead to the publication of guidelines to reduce visual manual distraction by the National Highway and Transportation Administration of the United States of America (NHTSA, 2012). Nowadays, the implementation of auditory-vocal systems has increased in order to minimize the distraction potential of the driver during the use of such information systems. Therefore, these systems require voice input and provide auditory feedback. This way, even though the direct visual-manual use of the device is successfully bypassed, distraction of the driver's attention may still occur through the cognitive workload involved.

In contrast to the large amount of research and measurement techniques that exist for the evaluation of visual-manual systems, research on cognitive workload is still in its early stages. Only a few widely accepted measurement techniques exist, which rely mainly on self reporting (e.g. De Waard, 1996). Therefore, the accurate evaluation of cognitive workload still constitutes a challenge of great importance for research.

In the following I will review briefly the state of research on cognitive workload, how it can be measured, and which quality criteria are therefore essential. Consequently, I will present an experiment that evaluates the sensitivity of a new evaluation method for the measurement of cognitive workload called detection response task (DRT) for use in the automotive context.

2. Theoretical background

2.1. Limited capacity

Driving requires a significant amount of cognitive workload in itself. This workload increases significantly through the use of various driver information and infotainment systems during the driving task, as the driver is required to divide his/her attention between the actual driving task and the use of these systems (Santos et al., 2005).

Attention is generally considered to be a limited capacity within the human information processing procedure and selective attention is therefore viewed as a logical consequence stemming from these limitations. With regard to the human information processing, it can be assumed that the availability of the resources required is indeed limited.

The concept of limited process resources within the human information processing can be found in numerous basic theories on attention (Broadbent, 1958; Kahnemann, 1973; Wickens, 1984). Distribution of those resources depends upon the tasks difficulty and the individual's motivation to successfully complete the task. Therefore, the ability to cope with demanding situations differs from individual to individual, as does the respective mental workload.

2.2. Resource theories

In the last decade, various plausible models of human resource management have been the focal point of the debate stemming from the need to understand the ways and forms in which cognitive resources may be compromised. Generally, two research traditions have been established by researchers, single resource theories and multiple resource theories which will be briefly described in the following sections.

2.2.1. Single-Resource-Theories

Single resource theories focus on a sole, central resource. According to Kahneman (1973) humans have access to a central pool of resources. If the available capacity of this single resource pool is exceeded (for instance as a result of multiple competing tasks) cognitive demand can be observed.

According to the single-resource theories, there is a direct connection between the number, the difficulty of the simultaneously attended tasks as well as the resulting cognitive performance: with additional tasks the performance deteriorates and influences the limited cognitive resource. The higher the difficulty level of a task, the lesser are the resources and as a consequence the worse is also performance.

This theory is usually considered in connection with dual task experiments. Two parallel tasks can be simultaneously performed, until the available single resource becomes exceeded and thereby is no longer sufficient to keep both tasks functional. When this occurs, deterioration of the performance may be observed, which mainly manifests in delayed reaction (for example delayed pressing on a key, or an increase of the latent reaction time before producing an oral answer) or faulty actions.

2.2.2. Multiple-Resource-Theories

Contrary to single-resource proponents, multiple resource theories suggest specific modules for information processing. Although this theory also suggests a limitation of the cognitive system, it determines the general capacity of the cognitive system as the combination of various single capacities, independent from one another.

The multiple resources model proposed by Wickens (2002) outlines four categorical dichotomous dimensions of information processing (see Figure 1):

a) **Stages**: perception & cognition vs. responding

b) **Codes**: spatial vs. verbal

c) **Modalities**: visual vs. auditory

d) **Visual processing**: focal vs. ambient

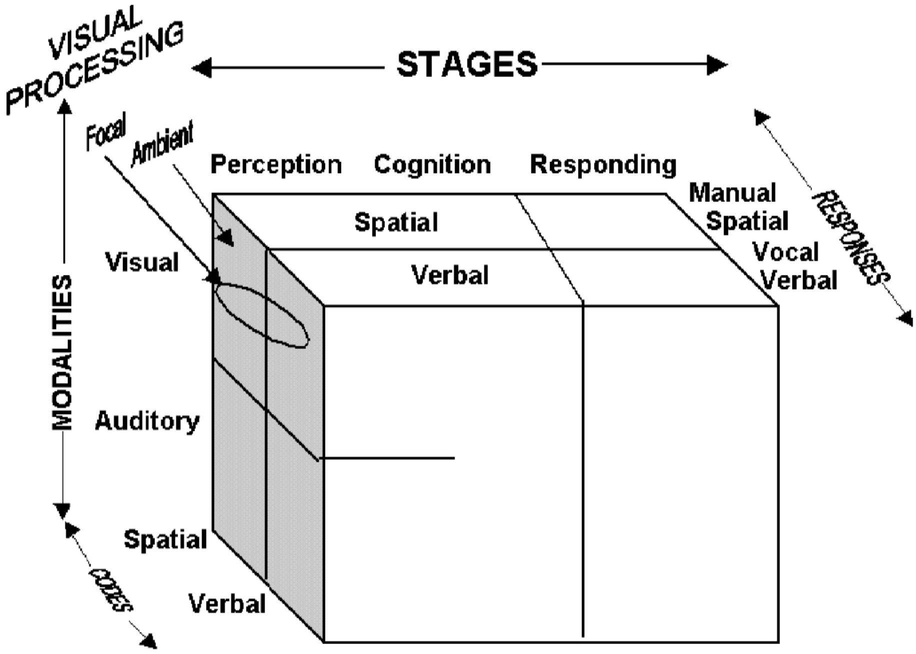

Figure 1 Structure of the Multiple Resource Model as proposed by Wickens (2002).

Wickens` model states that there are independent resources for perceptual and cognitive activities which are also separated from the underlying execution and response selection (see Figure 2). Evidence for this dichotomy is provided when the difficulty of responding in a task is varied and this manipulation does not affect performance of a concurrent task whose demands are more perceptual and cognitive in nature.

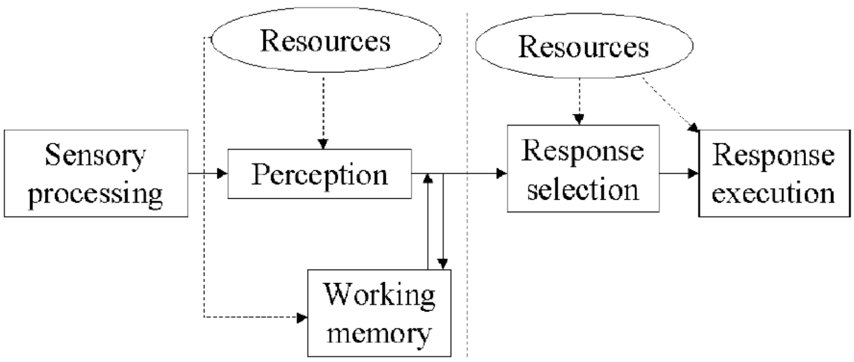

Figure 2 Representation of resources that supply different stages of information processing (Wickens, 2002).

The codes of the processing dimension indicate that spatial activity uses different resources than verbal/linguistic activity does, a dichotomy expressed in perception, working memory (Baddeley, 1986), and action (Liu & Wickens, 1992; Wickens & Liu, 1988). The separation of spatial and verbal resources accounts for the relatively high degree of efficiency with which manual and vocal responses can be time-shared.

The modalities dimension (nested within perception and not manifesting within cognition or response) indicates that auditory perception uses different resources than visual perception does and thus it is easier to divide attention between the eye and the ear than between the same channel (auditory or visual).

A fourth dimension was later added to these three dimensions: visual channels, distinguishing between focal and ambient vision, a nested dimension within visual resources. Focal vision, primarily (but not exclusively) foveal, supports object recognition and, in particular, high acuity perception. Ambient vision, distributed across the entire visual field and unlike focal vision preserving its competency in peripheral vision, is responsible for perception of orientation and movement. This can be explained with the help of the following example: ambient vision is needed to keep a car moving in the centre of the lane (e.g. Mourant & Rockwell, 1972), whereas focal vision is essential for reading road signs, glancing in the rear view mirror or recognizing hazardous objects on the road.

According to this model, it is possible to undertake multiple and parallel tasks without any undue disturbance if external conditions remain unchanged. For instance, the ability to divide attention regarding separate perception modalities (i.e. eye and ear) is easier by the so called cross-modal time-sharing than by the so called intra-modal time-sharing (Wickens, 2002).

According to both resource theories resources are limited. Thus, information processing is strongly connected to cognitive workload, which will be defined in the next chapter.

2.3. Definition of cognitive workload

Next to the discussion concerning which mental resources are available for the information processing, it is also important to define the term “workload”, or at least to attempt a clarification as to what the term means since the term has no standardized definition, even though the concept of workload might be familiar to most people intuitively.

A simple definition of workload would suggest that it is the sum of all demands, which have been placed upon an individual on a given moment in time. However, this definition defines workload only insofar as external influences are concerned (De Waard, 1996) which does not make any allowances for assessing how these influences affect an individual. Therefore, a more precise definition, which would cover both – the system requirements and the specific characteristics of the individual user - may be more appropriate. According to the definition proposed by Parasuraman and Hancock (2001):

> *„Workload may be driven by the task load imposed on human operators from external environmental sources but not deterministically so, because workload is also mediated by the individual response of human operators to the load and their skill levels, task management strategies, and other personal characteristics."*
>
> *(S.306).*

Accordingly, the degree of workload depends upon two components, on the complexity of the task (task load) and on the individual performance prerequisites (abilities, skills and motivational settings). Mental workload is thus the consequence of a complicated interaction between the specific traits of an individual and the requirements of the task in the context of motivation (Manzey, 1998).

Once both aspects have been incorporated in the definition, it becomes evident that the estimate of the workload depends on the nature of the task and the characteristics of each individual. It follows thus, that the same task requirements can induce distinctively more intense workload when applied to different people.

Different workload assessment techniques have been proposed. They will be discussed in the following chapter. However, I will first describe basic quality criteria for such techniques that will help to evaluate the quality of each workload assessment technique.

2.4. Workload-assessment techniques

The ability to correctly measure the driver's workload is essential for the development and evaluation of driver information and infotainment systems. The goal is to ensure that the workload involved in their use remains as low as possible in order to avoid potential perils such as automobile accidents.

2.4.1. Quality criterions for workload-assessment techniques

Since workload is not directly observable, measurement methods especially designed for this purpose need to be developed. Evidently, these methods have to comply with specific quality criterions. According to O`Donnell and Eggemeier (1986) for a more precise workload assessment technique additional criteria such as sensitivity, diagnosticity, primary task intrusion, operator acceptance and implementation requirements need to be taken into consideration alongside usual psychological quality criteria such as objectivity, validity and reliability that will be described below.

Objectivity: A measurement fulfills the objectivity criterion of quality when two distinct observers acquire the same result when equipped with the same measuring instrument (for the same objective).

Validity: The validity of a measuring instrument depends on the extent to which the instrument in fact measures the dimension it claims to measure. In this case this dimension refers to the extent to which the method reacts exclusively with variations of cognitive workload.

Reliability: This dimension concerns the reliability of a method. This means its stability and consistency. A method is therefore reliable if it produces consistent results under consistent conditions.

Sensitivity: This criterion describes the capability of the technique to indicate changes of the workload level due to task difficulty or resource demand. The more accurately mental workload fluctuation can be registered, the more sensitive the technique is considered to be.

Diagnosticity: The term diagnosticity refers to a method's potential to reflect demands on a specific resource, i.e. to what extent the method's procedure can provide with information concerning the underlying factors influencing cognitive workload.

Primary-task intrusion: This dimension refers to the degree to which a workload assessment technique interferes with the primary task. As the primary task in determining the degree of driver distraction is driving, the applied measurement tool should not degrade the driving performance.

Operator Acceptance. The degree of approval of the technique by the user is referred to as operator acceptance. The user's opinion of the measurement technique can affect the correctness and accuracy of the measure. In general the acceptance is higher if the technique is less intrusive and has high face validity.

Implementation Requirements: This dimension assesses the practical limitations of a technique, such as, for example, the requirement of a specific piece of equipment or possession of specialized technical knowledge.

For the following section established measuring methods for cognitive workload will be introduced and discussed. According to O'Donnel and Eggemeier (1986), the empirical measuring techniques can be classified in three distinct categories:

- self-report measures
- physiological measures
- performance measures

2.4.2. Self-report measures

For this type of evaluation method, the user is questioned on the degree of his/her mental workload as well as on the way upon which this manifests. Subjective, self-report measures are based upon the premise that the user is capable to correctly identify and assess the degree of his/her mental workload, and the limitations of his/her own capacity to process information (Rößger, 1996).

Indexing is normally done with the use of a single or multi dimensional rating scale. The following are the most prominent examples of such subjective measures:

- Rating Scale of Mental Effort (RSME) by Zijlstra (1993)
- NASA-TLX (NASA Task Load Index) by Hart and Staveland (1988)

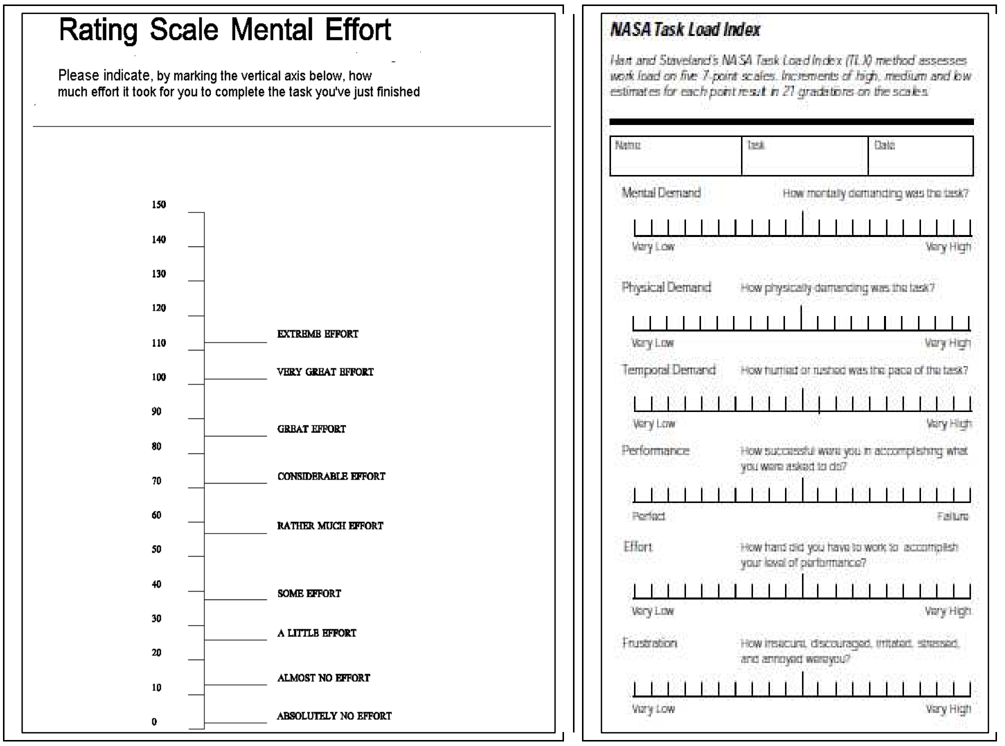

Figure 3. Examples of the RSME and NASA-TLX surveys.

The single-dimensional RSME offers questions about the subjective effort involved in solving a task, whereas the multidimensional NASA-TLX measures mental and physical workload by means of six dimensions (mental demand, physical demand, temporal demand, performance, effort, frustration).

In most cases, the survey is conducted directly after task processing. Questioning during task completion has been deemed problematic since, according to Wickens and Hollands (2000), the number of tasks which need to be executed simultaneously is essential at the assessment of workload. This means that two easy tasks, which can be executed simultaneously without any significant problems could be assessed as more difficult than a single difficult task which fails. This could lead to adverse effects.

Furthermore, to define workload status on a given moment through subjective means is impossible, since conducting the questioning momentarily interrupts the process and can therefore lead to erroneous measurements.

Moreover, a subjective measurement of the degree of cognitive workload is based on information recalled at the time of the survey and which respectively the subject has become consciously aware of (Hart & Staveland, 1988).

2.4.3. Physiological measures

The second option available for measuring cognitive workload is to collect physiological data. The fundamental idea behind this is that changes of the workload can be directly observable in changes of the central or the vegetative nervous system. Typical indications would include, for instance, variation of the heart frequency, skin resistivity and EEG.

A significant advantage of the measuring of mental workload by means of physiological examination is that it does not require participation of the user in a direct way and measurements may be obtained during task oriented activities, without interrupting the process. Moreover, physiological measurements are more concise and offer a higher temporal resolution (Wickens & Hollands, 2000).

However, these measuring methods are heavily dependent on specialized technical equipment, as the case is, for example, for the electroencephalogram (EEG). This method is currently not practically applicable for the evaluation of in-vehicle driver distraction. Another recently discussed method is the so called pupillometriy (Schwalm et al., 2008).

Pupillometry is based upon the observation that pupils do not react only to stimulation by light, but also to emotional and mental processes, and that the size of the pupil is altered depending on the effects of mental workload.

The equipment required for such measurements is a highly sensitive eye tracking system, which can register changes of the pupil diameter with high temporal frequency. The high resolution of the measurement gives this procedure a significant advantage over other methods, as the continuous change in mental workload can be observed.

Nevertheless, such a measurement is not entirely problem free, since light conditions and changes in the distance can also trigger alterations of the pupil diameter. Therefore, if applied to realistic and application oriented situations, like in a driving experiment, the challenge lies in correcting the raw pupil signal by eliminating external influences. This can be done by using a special method called the “Index of Cognitive Activity” (Marshall et al., 2004).

Although this method seems to be very suitable to measure mental workload while driving, it is in its current stage still a more research oriented approach. The underlying algorithms are still protected by a patent of a specific company. Further, the method has only been used in few research studies. Therefore, validation of the method and its algorithms is still research in progress (Schwalm, 2009).

2.4.4. Performance measures

The third variant of empirical measuring techniques is to measure performance. According to this model, behavior or performance during task procedure are evaluated and measured so that the workload involved may be determined. Performance measurements can be further distinguished into two sub categories: primary task measures and secondary task measures.

Primary task measures

A task is considered primary, when it is aimed towards the main focus of attention. Such primary activities are ascribed the highest degree of priority. Performance measurement is very task specific, given that the used primary tasks can vary considerably. The number of errors committed, the speed of performance or the reaction time measures are frequently used as primary-task performance measures in laboratory tasks. There is not one prevalent primary task measure, although all primary-task measures are speed or accuracy measures. In relation to driving typical performance measures are speed, lane keeping (lateral and longitudinal control) and steering angle.

However, there is a considerable disadvantage in the use of performance measures. They can not provide with an indication as to what the actual mental workload of the operator might be. In addition it can fail to detect performance differences between two individuals, as one may be approaching the limit of his/her capacity while the other may still have available resources at his/her disposal (De Waard, 1996).

Therefore it is necessary to combine primary task performance with other dependable workload measures in order to draw valid conclusions about man-machine interaction and, in particular, about the operator's strategy or energetic state.

Secondary task measures

Performance measure of secondary tasks is another way to evaluate mental workload. In such scenarios, a secondary task is being performed alongside the primary task and the performance exhibited in this additional task is evaluated (O`Donnell & Eggemeier, 1986).

This is based upon the theoretical suggestion that all resources not otherwise engaged by the performance of the primary task are still available, and may be used for other activities, which relate to the secondary task. Secondary task performance evaluation allows for conclusions to be drawn on the extent of the use of resources and the workload generated by the primary task.

According to this approach, the Peripheral Detection Response Task (PDRT) was introduced by Martens and Van Winsum (2000), to use performance in a secondary task as a gauge in order to determine cognitive workload. The PDRT constitutes the first variant of Detection Response Tasks (DRTs). In my experiment I will compare three variants of this new workload assessment technique.

2.5. Detection Response Task

2.5.1. Development of the Peripheral Detection Response Task

The Peripheral Detection Response Task (PDRT) was originally developed by van Winsum, Martens and Herland (1999) at the TNO Human Factors Research Institute in the Netherlands. The PDRT is based on the perception of visual stimuli, which are presented in the periphery of the visual perception range. Responses are typically made by pressing a microswitch attached to the index finger. Thereby the driver’s cognitive workload is assessed indirectly via the detection rate. The development of this novel evaluation method was based on the assumption that the size of the field of perception is reduced through the application of increased cognitive workload and that a so called visual tunneling effect occurs.

This idea can be traced back to Miura (1986), who presented spots of light on the windscreen under different horizontal angles and measured the reaction time for stimuli detection while conducting a driving experiment. He found that the reaction time increased as the complexity of the driving task (e.g. higher traffic density) simultaneously intensified. The results were interpreted as being indicative for a reduction of the functional visual field of view when the complexity of the driving task becomes higher.

Similar results have been reported by Williams (1985, 1988), who conducted tachistoscope studies, where the gaze direction was fixed. Participants were asked to name letters as a primary (foveal) task and identify line orientation as a secondary (peripheral) task. He noted an interaction between eccentricity and foveal task load, albeit only in some conditions, which can be interpreted as tunnel vision. The ability to process peripheral information decreased as foveal load increased.

However, Recarte and Nunes (2003) investigated if this effect is due to visual tunneling or if it is in fact caused by general interference. They tested the effect on detection and discrimination of visual stimuli with different eccentricities and found significant effects of cognitive load but no interaction with eccentricity. This clearly provides indication that the cause of the decreased performance lies with a general interference rather than visual tunneling.

In fact, “cognitive tunneling effect” suggested by Viktor et al. (2008) may be a more appropriate term to use, when it comes to describing the phenomenon than “visual tunneling effect”, given that the phenomenon is indicative of a shift towards increasingly selective patterns of attention.

In their study Martens and van Winsum (2000) tested the PDRT with stimuli at different eccentricities in high and low driving demand conditions. They found no evidence for a visual tunneling effect, as there was no effect of visual eccentricity of the stimuli at all. (See Figure 4).

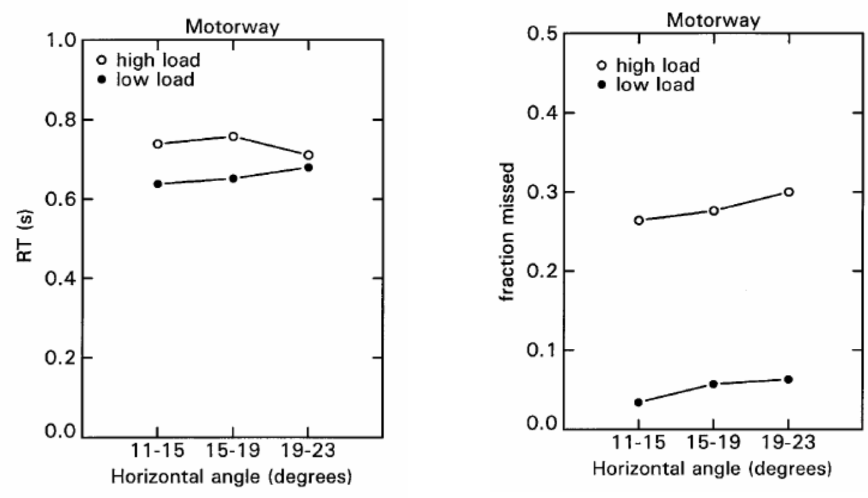

Figure 4. RT and fraction of missed signals as a function of horizontal angle and workload (Van Winsum et al., 1999).

Thus, it can be asserted that it is rather a cognitive tunneling effect that occurs, and that the detection performance depends surprisingly little on the stimuli position. Moreover, the sensitivity of the PDRT does not depend on the stimuli position at all.

The main application of a detection response task in research on cognitive workload until now has been the visual PDRT. The PDRT has been used in simple laboratory studies (e.g. Liu et al., 2009) as well as in simulator studies (e.g. Burns et al., 2000) as well as in real traffic (e.g. Jahn et al, 2005). In all of these applications the PDRT has since been established as a highly sensitive method for measuring cognitive workload (e.g. Olsson & Burns, 2000; Harms & Patten, 2003; Jahn et al., 2005).

However, there are some clear drawbacks and limitations of the classic PDRT. The PDRT depends on the participants` ability to see the stimulus and at least in non-laboratory studies it can not be excluded that head and eye movement have an impact on stimulus detection. As an example of such a negative impact it was shown by Jahn et al (2005) that participants missed many PDRT signals while waiting at traffic lights.

To summarize, while laboratory studies showed no effects of stimuli position, there was a clear negative impact of traffic situation in field studies using the PDRT. To overcome such inherent methodological drawbacks of the classic PDRT, new research has focused on establishing other versions of the DRT as well as discussing the suitability of such DRTs for laboratory as well as field studies (Engström, 2010).

2.5.2. Different Detection Response Tasks

Since research showed, that the sensitivity of the PDRT does not depend upon the position of the stimuli presented, a next logical development was to explore whether it is in fact essentially independent from the stimulus modality.

Therefore a number of modifications such as the Auditory Detection Response Task (ADRT) and the Tactile Detection Response Task (TDRT) have been developed. The stimuli involved in these modifications are not visual, but auditory and tactile correspondingly, eliminating the problem of looking away and thereby missing visual stimuli.

All three DRT variants have been (simultaneously) tested for the first time in a single driving simulator study by Merat and Jamson in 2008. In their experiment, they examined the effect of two tasks on signal detection in the visual, auditory and tactile modalities. The tasks used were one visual-manual telephone task (typing a seven digit phone numbers on a touchscreen numberpad) and a cognitive task, where participants had to count backwards. The results showed that all three types of DRT have the same sensitivity and that all three detection tasks are suitable for assessing distraction (see Figure 5).

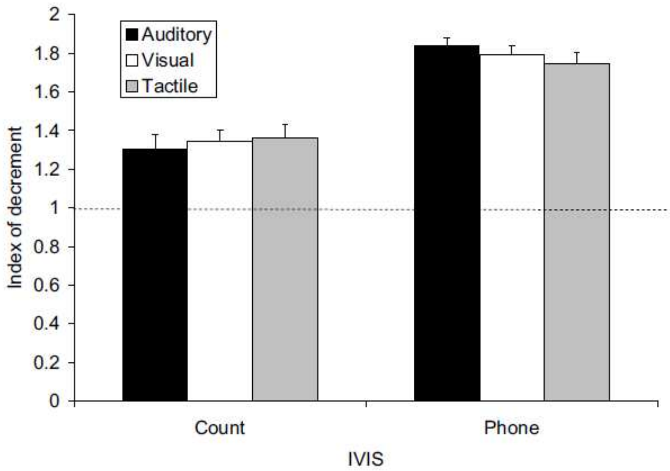

Figure 5. The effect of each IVIS on index of decrement in reaction time for each detection task (Merat & Jamson, 2008).

While the authors of this study argue that their results show that all three types of detection tasks and their corresponding modality can be exchanged according to the test conditions (e.g. replacing LED with auditory stimuli on a sunny day) it is not yet clear to what degree their results generalize.

Firstly, they used a tertiary design (driving + test task + DRT) as experimental set-up. Driving scenarios vary to a large degree as well as driving simulators do. To compare results between different test sites, one would have to standardize a driving simulation scenario that had to be used along with the DRT. However, such standardizations of driving simulators and scenarios have proven very difficult (Jamson, 2000). Furthermore, such tertiary designs can add complexity to the evaluation procedure. Therefore, it would be worth to evaluate different DRT tasks in a more basic setting as a secondary task (test task + DRT).

Secondly, Merat and Jamson used one visual-manual task and only one cognitive task. They did not vary experimentally how such cognitive tasks are presented or induced (e.g. peripheral, auditory, or pure cognitive) and failed to account for interactions between the modality of the detection task and the modality of the test task as predicted by Wickens` model (see Figure 1).

For this reason the following experiment is based on a more fundamental, basic research approach. The cognitive workload was systematically varied using two difficulty levels as well as different presentation modalities of the cognitive task to determine if interferences between the DRT variants and the presentation mode of the secondary task exist.

3. Research Objective

For the purposes of this study three different DRT variants (peripheral, auditory and tactile) are systematically tested in reference to their sensitivity to measure cognitive workload under different presentation modes of the secondary task.

According to Wickens` multiple resource theory, dual task interference will be more intense when two tasks demand overlapping resources. It is therefore important to determine whether the chosen type of DRT applied has any influence on the secondary task. In particular whether it triggers undesirable interaction when the stimuli are presented via the same channel (visual/auditory) on which also the secondary task itself is being loaded.

As this work aspires to contribute to the fundamental research into the different DRTs, artificial test tasks were chosen. In order to test the effect of high and low imposed cognitive workload on the detection response task performance, cognitive loading secondary tasks were applied. These additional tasks were implemented with two levels of difficulty: Easy, i.e. less demanding task, and difficult, i.e. more complex and thus more demanding task, in order to verify that the DRTs are an appropriate measurement tool by which cognitive workload can be assessed effectively.

Finally the following secondary tasks were chosen: The n-back task (Mehler, Reimer, Coughlin & Dusek, 2009) via visual presentation adapted from the original requirements and additionally in the original version via auditory presentation. A counting task, designed to serve as an almost pure cognitive presentation mode, was also implemented. Even though these tasks differ in their presentation mode, they all induce comparable low and high cognitive workload.

It becomes therefore important to determine whether an interference occurs as suggested by the multiple resource model by Wickens (2000). For instance, the peripheral detection task (PDRT) and the visually presented n-back task use the same resources and thus the performance might be negatively affected. The same hypothesis is extended to include both the auditory detection task (ADRT) and the aurally presented n-back task. As a control the counting task should serve as a pure cognitive workload inducing task and therefore no interferences with any of the DRT variants were expected here.

4. Method

4.1. Participants

All twentyfour participants (12 female, 12 male) in this experiment worked for BMW. The age range of the subjects tested was between 21 and 42 years old, with an average of 29 years (*SD* = 5.17). All of the participants had normal or corrected-to-normal vision. Three of the participants were left handed, but one of them was forced to use his right hand as a child and is now ambidextrous.

4.2. Apparatus

The experiment was carried out in the Usability Lab 2 at the Research and Innovation Center of BMW in Munich. The experimental set-up was built upon a desk and the participants were asked to assume a centric position in front of it (see Figure 6).

The experimental set-up consisted of a laptop, two loudspeakers, positioned behind the laptop and the three different detection response tasks:

Figure 6. Experimental set-up.

For the peripheral detection response task (PDRT) five red LEDs were mounted horizontally on a black cardboard and spread symmetrically from the center point. The participants were seated 1 m in front of the LED bar so that the stimulus presentation would be obtained from a horizontal angle of 11° to 23° of the participants` forward view, as specified by Martens and van Winsum (2000). Only the four outer LEDs were lit up, the one in the middle was used to measure the correct distance.

Participants were required to respond as soon as they detected a lit up LED by pressing a microswitch, which was attached to the index finger of their dominant hand. On average every 4 s, with a random variation between 3 and 5 s, a visual stimulus was presented. The LED signal was visible for a maximum of 1 s and within this limited time frame it disappeared as soon as the subject gave a response.

The auditory detection response task (ADRT) was realized by presenting a 1 kWh sinus tone (according to Merat & Jamson, 2008) using two loudspeakers arranged behind the laptop. The participants were asked to adjust the volume to a comfortable audible level prior to the start of the experiment. The auditory stimuli were produced for a maximum duration of 1 s and on average every 4 s, with a random variation of 3 and 5 s. Prior to the initiation of the experiment, participantss had been instructed to press the microswitch upon detection of an auditory stimulus which caused the sinus tone to stop.

Finally, a small electrical vibrator, obtained from a mobile phone (according to Engströn et al., 2005), was attached to the wrist of the non dominant hand of the participants (see Figure 7) so that testing for the tactile response task (TDRT) may be conducted. The participants were encouraged to adjust the vibration strength via a regulator to a comfortable level prior to the start of the experiment. The tactile stimuli were also given every 4 s, with a random variation between 3 and 5s and lasting a maximum of 1 s. The subjects were asked to respond as soon as they detected a tactile stimulus by pressing the microswitch.

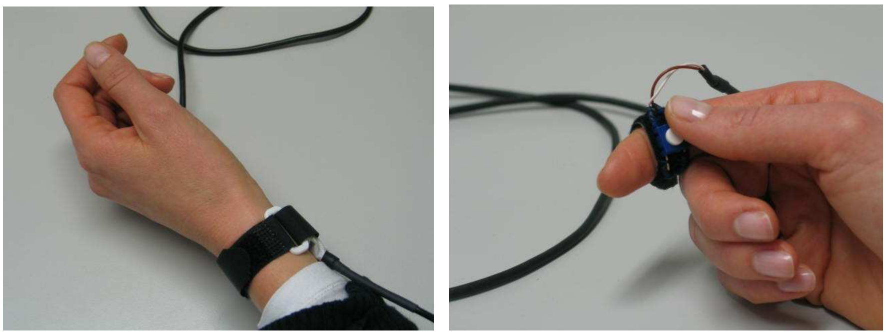

Figure 7. Vibrator pad attached to the wrist and microswitch.

Average reaction time and hit rate were used as performance indices for all three variants of the DRTs. Responses that were given within 2 s after the onset of a detection response signal were counted as a hit. The hit rate therefore corresponds to the percentage of correctly responded stimuli within the 2s time frame. Any subsequent response outside the 2 s time frame following the launch of the stimulus was considered as a miss. The complete lack of a response has also been registered as a miss and has been included in the calculation of the rate. Additional responses triggered in the absence of a stimulus were considered as false alarms.

4.3. Secondary Tasks

For the auditory presentation mode the n-back task as specified by Reimer (2009) was used. 10 single-digit numbers (0 to 9) with an interstimulus interval of 2.5 s were recited aurally via loudspeaker.

Two levels of difficulty were employed; the 0-back and the 2-back version. In the 0-back version the participants were asked to simply repeat out loud each number immediately after it was presented (see Table 1 for illustration). This variant is considered easy and imposes less cognitive workload, unlike the 2-back version which is highly demanding. In this condition the participant is required to recall from memory the number that was presented two numbers before the current value (i.e. two items back) and repeat out loud that value meanwhile the next numbers are presented. This places the 2-back version in the 'difficult' category of variants.

The following example illustrates in detail how the 0- and 2-back task were implemented:

Table 1. Example of the two versions of the n-back task.

Presentation:	0	4	7	1	3	9	2	8	5	6
0-back	0	4	7	1	3	9	2	8	5	6
2-back	Subject is silent	Subject is silent	0	4	7	1	3	9	2	8

The n-back task was also presented visually via PowerPoint presentation on a laptop to cover the visual presentation mode. The original aural set-up was hence transformed into a visual presentation mode with all above mentioned specifications remaining the same. In order to eliminate the learning effect, new series of numbers were employed. Again the two levels of difficulty (0- and 2-back) were used.

As a third task without any real presentation mode the counting task was implemented. The participants were requested to begin counting upwards in steps of two beginning from a given three-digit number. This was the easy variant. More workload was induced by the difficult variant, where the participants were asked to count downwards in increments of seven from a given three-digit number. See Appendix 9.1 for the n-back audio files and PowerPoint presentations, as well as the given three-digit numbers.

4.4. Design

In general a 3x2 within-subjects design was employed for each detection response task, with three different presentation modes (visual, auditory, purely cognitive) and two levels of difficulty (easy and difficult). The dependent variables were reaction time and hit rate.

4.5. Procedure

The experimenter offered all participants a brief familiarization period and adequate explanations both on the detection response tasks and on how the various devices and apparatuses worked prior to experimentation. The participants were able to adjust the volume of the auditory stimulus for the auditory detection response task, as well as the strength of the tactile stimulus administered by the vibrating element attached to their wrist to a comfortable level.

Once the initial stages of the experiment were completed, the secondary tasks were introduced and explained. Even though it was not possible for the participants to train for them, adequate instructions and explanations were provided about what was required of them and how the procedure functioned. Additionally, the participants were instructed to not prioritize any one of the two given tasks (detection response task and secondary task) in favor of the other, but rather to aim for the fastest and most accurate performance possible of both tasks.

The setting of the detection response tasks was carried out blockwise. The participant started the experiment with one variant of the DRT and ran through all the three secondary tasks, proceeded with the next DRT variant and finished with the last one. The order of the DRT variants and the secondary tasks was randomized. Each secondary task was repeated three times, whereas the first trial served as training and its results were not used for subsequent analysis upon completion. The total testing time lasted approximately one hour.

5. Results

The aim of the study was primarily to determine whether the DRTs are an effective means to measure cognitive workload, and as such able to clearly differentiate between less and more demanding tasks, and secondly to determine which of the tested DRTs is best suited to which mode of presentation.

Analyses of variance (ANOVA) were carried out on the current data set to determine if there were any effects of difficulty for each secondary task in the reaction time and hit rate. T-tests were run for each DRT variant to ascertain whether they are discriminating between the low and high imposed cognitive workload.

In Figure 8 the mean reaction times as well as the mean hit rates are shown for the visually presented n-back task.

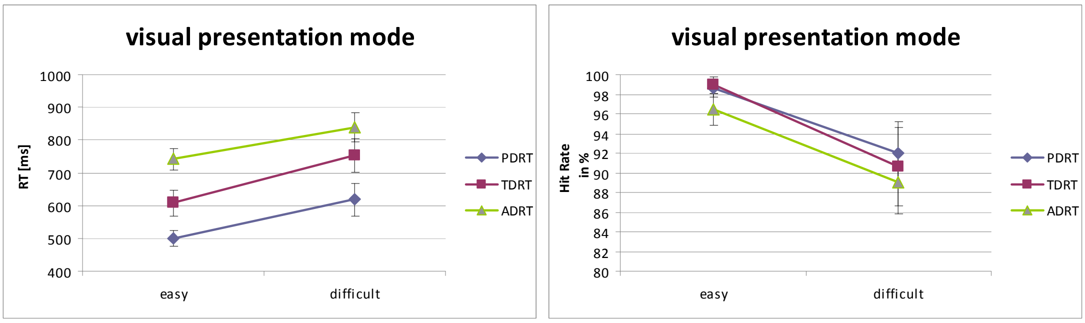

Figure 8. Mean RTs and hit rates for the three different DRT variants under the visual presentation mode.

For the visual presentation mode, significant main effects of the difficulty were found for the reaction time, $F(1,23) = 32.68$, $p < .05$ and the hit rate $F(1,23) = 13.48$, $p < .05$.

T-tests for the three different DRT variants revealed that all of the variants proved to be effective and able to differentiate between low and high cognitive workload. (Reaction time: PDRT $t(23) = -3.15$, $p < .05$, $r = 0.55$; TDRT $t(23) = -4.72$, $p < .05$, $r = 0.70$; ADRT $t(23) = -3.41$, $p < .05$, $r = 0.58$; hit rate: PDRT $t(23) = 2.31$, $p < .05$, $r = 0.43$; TDRT $t(23) = 2.90$, $p < .05$, $r = 0.52$; ADRT $t(23) = 3.23$, $p < .05$, $r = 0.56$).

Table 2 shows the mean reaction times and hit rates for the different DRT variants in the visual presentation mode.

Table 2. Mean RTs (ms) and hit rates (%) plus SD for different DRT variants.

		PDRT easy	**PDRT** difficult	**TDRT** easy	**TDRT** difficult	**ADRT** easy	**ADRT** difficult
Reaction Time	M	501	618	608	754	741	840
	SD	116	240	181	208	157	216
Hit Rate	M	99	92	99	91	96	89
	SD	3	14	3	16	7	13

Figure 9 shows the average reaction times as well as the mean hit rates for the auditory presented n-back task.

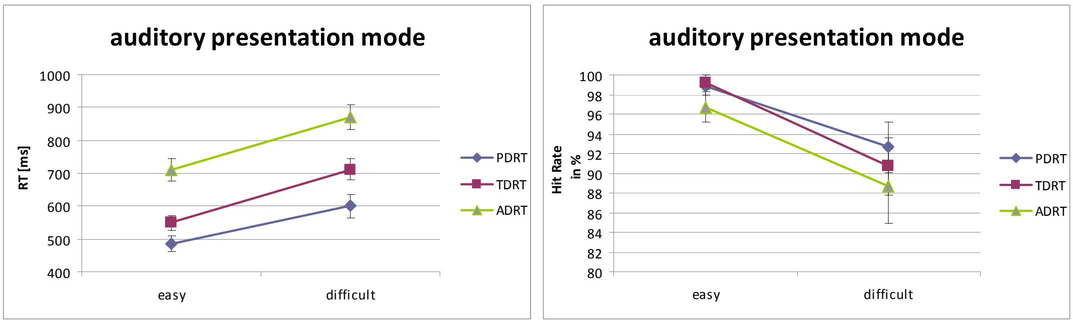

Figure 9. Mean RTs and hit rates for the three different DRT variants under the auditory presentation mode.

Significant main effects of the difficulty were also found for the reaction time, $F(1,23) = 129.45$, $p < .05$ and the hit rate $F(1,23) = 19.55$, $p < .05$ under the auditory presentation mode.

T-tests for the three different DRT variants revealed that all three were able to significantly differentiate between low and high cognitive workload to a significant degree.

(Reaction time: PDRT $t(23) = -6.81$, $p < .05$, $r = 0.82$; TDRT $t(23) = -6.67$, $p < .05$, $r = 0.81$; ADRT $t(23) = -9.27$, $p < .05$, $r = 0.89$; hit rate: PDRT $t(23) = 3.34$, $p < .05$, $r = 0.57$; TDRT $t(23) = 4.13$, $p < .05$, $r = 0.65$; ADRT $t(23) = 2.45$, $p < .05$, $r = 0.46$).

The mean reaction times and hit rates for the different DRT variants in the auditory presentation mode are shown in Table 3.

Table 3. Mean RTs (ms) and hit rates (%) plus SD for different DRT variants.

		PDRT easy	**PDRT** difficult	**TDRT** easy	**TDRT** difficult	**ADRT** easy	**ADRT** difficult
Reaction	M	486	600	549	712	711	870
Time[ms]	SD	114	171	107	164	166	183
Hit	M	99	93	99	91	97	89
Rate [%]	SD	3	9	3	11	5	16

Once again, all of the three DRT variants demonstrate the capacity to significantly discriminate between low and high cognitive workload that is orally presented.

In Figure 10 the mean reaction times as well as the mean hit rates are shown for the counting task.

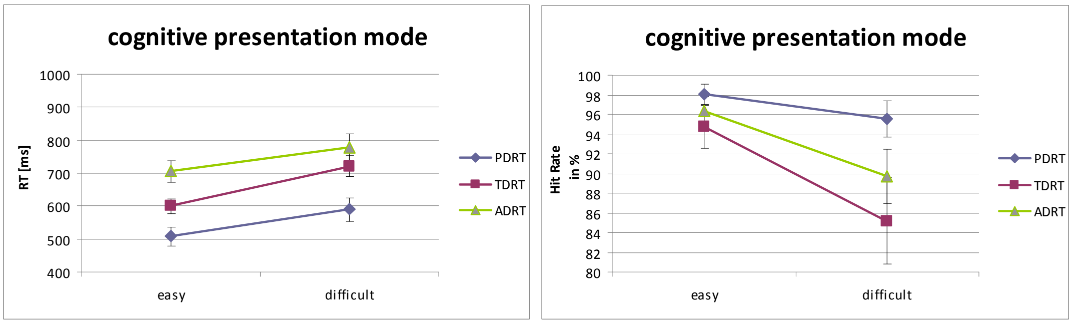

Figure 10. Mean RTs and hit rates for the three different presentation modes.

For the purely cognitive presentation mode, a significant main effect for difficulty was found for the reaction time, $F(1,23) = 29.15$, $p < .05$ and the hit rate $F(1,23) = 18.44$, $p < .05$. However, there was also a notable interaction for the hit rate between the difficulty and the DRT variants, $F(2,46) = 3.3815$, $p < .05$ indicating that the different DRT variants exhibit varying degrees of sensibility when the difficulty level of the task increases.

T-tests for the three different DRT variants revealed that all of them are able to significantly discriminate between low and high cognitive workload regarding the reaction times. (Reaction time: PDRT $t(23) = -3.25$, $p < .05$, $r = 0.56$; TDRT $t(23) = -3.44$, $p < .05$, $r = 0.58$; ADRT $t(23) = -2.63$, $p < .05$, $r = 0.48$)

However for the hit rate only the TDRT and the ADRT are a sensitive measure. (Hit rate: PDRT $t(23) = 1.92$, $p > .05$, $r = 0.37$; TDRT $t(23) = 3.45$, $p < .05$, $r = 0.58$; ADRT $t(23) = 3.18$, $p < .05$, $r = 0.55$)

Table 4 shows the mean reaction times and hit rates for the different DRT variants in the cognitive presentation mode.

Table 4. Mean RTs (ms) and hit rates (%) plus SD for different DRT variants.

		PDRT easy	**PDRT** difficult	**TDRT** easy	**TDRT** difficult	**ADRT** easy	**ADRT** difficult
Reaction Time	M	508	589	600	722	705	777
	SD	143	174	188	253	158	207
Hit Rate	M	98	96	95	85	96	89
	SD	3	6	8	16	4	10

Thus, it is evident from the results presented here that there were no interferences between the DRT variant and the presentation mode at all. The results are almost identical for each presentation mode.

6. General Discussion and Conclusion

Recently a shift towards the implementation of auditory-vocal information systems (NHTSA, 2012) into the vehicle can be observed, which aimed to reduce the distraction caused by visual-manual information systems. Nevertheless, these systems may be still distracting as cognitive workload occurs. To assess the cognitive workload that is imposed by these systems, suitable measurement techniques are needed. Therefore the purpose of this study was to take first steps in order to evaluate new types of cognitive workload assessment techniques: detection response tasks (DRTs). Three different types of DRTs were compared to evaluate their suitability to measure cognitive workload.

The following DRT variants were evaluated: peripheral detection response task (PDRT), auditory detection response task (ADRT) and tactile detection response task (TDRT). The cognitive workload was systematically varied, by using artificial secondary tasks in two levels of difficulty which were presented visually, audibly and purely cognitive.

Thus it was determined if there are interferences between the DRT variant and the presentation mode of the secondary task both are presented via the same channel, e.g. visually.

The results of this study strongly suggest that all DRT variants were able to discriminate between high and low induced cognitive workload. The picture over all secondary tasks remains consistent while evaluating the reaction times. The reaction times in the PDRT setting are the shortest, followed by the reaction times with the TDRT and the ADRT setting. All three detection response tasks show a remarkable sensibility to measure cognitive workload.

Nonetheless it may be important to note that this picture may appear rather different should evaluation be based solely on the hit rate instead. A significant interaction between the detection response variant and the difficulty level indicates that the DRT variants do not demonstrate the same sensitivity when the level of difficulty increases. A closer examination reveals that the PDRT is not adequately sensitive to measure increased cognitive workload on the counting task. Thus it seems that the hit rate is an even more sensitive measurement than the reaction time alone.

Surprisingly no interferences between the modality of the secondary task and the DRT variants were detected. Even though the secondary task and the stimuli of the detection response task were sent to and received through the same channel (visual or auditory), there appeared to be no evidence of interference. This finding is contrary to the assumption and prediction based on the multiple resource theory by Wickens (2002). Where the modality of perception distinguishes between auditory and visual resources and therefore dividing attention between the ear and the eye should be easier than between the same channel.

A possible explanation for this might be that the DRTs themselves require so little in attentional resources. This of course is a useful discovery and generally speaking a positive one, considering the workload measurement tool should not induce a lot of workload itself. Therefore, all DRT variants are equally suitable for what ever kind of presentation modality of the secondary task at least with these artifical test tasks. However, another explanation could be that the cognitive tasks that were chosen were not as demanding and therefore participants did not reach their information processing limits.

Relating to the basic methodological quality criteria it can be stated that the DRTs are a valid measuring tool to assess cognitive workload. This is clearly shown in their ability to detect and discriminate between high and low levels of cognitive workload. They are furthermore a very objective technique, as the procedure and the interpretation of the results are independent of the experimenter. The reliability of the DRTs has to be more thoroughly examined in the future, as up until now no identical studies have been carried out. Although a consistent coherence between the DRTs and the cognitive workload in other studies could be found (Martens & van Winsum, 2000; Harms & Patten, 2003; Jahn et al., 2005; Patten et al., 2006).

The further quality criteria postulated by O`Donnell and Eggemeier (1986) by which the suitability to measure cognitive workload is assessed, are partly met. All three types of the DRT are very sensitive in their ability to discriminate between the high and low cognitive workload that was induced on the participants. However, the DRTs are very low in their diagnosticity, as they do not discriminate between the different types of workload. On the basis of DRT measures, it is not possible to distinguish between different types of workload (for example visual and cognitive workload). Since an adequate measuring tool for workload was needed though, this criterion is certainly not the most important one. Primary-task intrusion will always arise as soon as a secondary task is deployed. Nevertheless, the results of this study showed that the interference between the DRT variant and the presentation mode of the secondary task was negligible. As the operator`s acceptance of the assessment technique might have an influence on the measurement (e.g. low motivation) the DRT is quite suitable, because usually it is not recognized as a measurement technique itself, but rather as just another task. Finally the implementation requirements are rather low, compared to the pupillometry method for example. The different variants are easy to implement and data collection can be fully automated. Furthermore the cost of the hardware is negligible and the equipment relatively easy to set up.

Despite the fact that there is a strong consensus that the DRTs are indeed a very sensitive method of measurement for cognitive workload (Van der Horst & Martens, 2010; Engström, 2010), future research is still needed in order to specify some absolute criterion against which driver distraction can be more accurately determined, particularly in the context of international standardization.

As Olsson and Burns (2000) suggested permissible PDRT hit rates no less than 65% and reaction times not slower than 800ms as tangible thresholds. Admittedly until now there is no indication of how much impairment in these detection tasks is considered to be too much impairment. However, previous studies indicated a rise in the brake reaction time of 168 ms by a blood alcohol concentration that varied around the legal limit (e.g. de Waard & Brookhuis, 1991). Thus, it should be considered, if the use of any device that increases the reaction time of the driver by more than that of the legal alcohol blood concentration is advisable.

This study had its main focus on the detection of cognitive workload induced by artificial tasks, as it should provide further insight in basic research concerning the DRTs` sensibility. Therefore, to be able to draw a final conclusion about the suitability of the DRTs as an appropriate cognitive workload measuring tool, it is needed to test real world tasks in a similar systematical way. Such real world tasks like an automotive speech based in-vehicle system could show interferences with the ADRT that were not observed in this study. Accordingly, a real world visual manual task interference with the PDRT could be observed. These interferences could potentially come about as a result of higher levels of task complexity and workload.

Finally, an open question remains on how the DRTs should be applied if used to evaluate cognitive workload in an automotive context. Further testing is required in order to determine the extend to which it is possible to apply one of the DRT variants to a triple test scenario (driving + secondary task + DRT variant). The triple test scenario variant is clearly the one with highest face and ecological validity, because a driving scenario is combined with the test scenario. However, it is also much more complex and interferences between the three tasks can not be eliminated. Moreover it is to be expected that participants will use different compensation strategies to reduce the applied workload in especially demanding situations, for example by reducing speed (HASTE, 2004). Another possibility to reduce workload is that participants switch from triple task to dual task performance by neglecting the detection response task. Therefore the hit rate is an essential marker for the quality of the data. However it is crucial to use these tertiary designs as they offer high universal validity. Only the findings of using an information system while driving can be generalized to its real distraction potential.

According to the results of this study the use of the TDRT and ADRT appears to be the most appropriate and can be recommended, as for both the reaction time and the hit rate showed their sensitivity to measure cognitive workload. However, more empirical and theoretical work is needed, to gain more insight in the underlying concepts and to establish the DRTs as a valid method to evaluate cognitive workload respectively driver distraction.

7. References

Angell, L., Auflick, J., Austria, P. A., Kochhar, D., Tijerina, L., Biever, W., Diptiman, T., Hogsett, J., & Kiger, S. (2006). *Driver Workload Metrics Task 2 Final Report*. (NHTSA Report No. DOT HS 810 635) Retrieved from NHTSA website: http://www.nhtsa.gov/DOT/NHTSA/NRD/Multimedia/PDFs/Crash%20Avoidance/Driver%20Distraction/Driver%20Workload%20Metrics%20Final%20Report.pdf

Baddeley, A. D. (1986). *Working Memory*. Oxford: Oxford University Press.

Broadbent, D. E. (1958). *Perception and communication*. New York: Pergamon Press.

Burns, P. C., Knabe, E., & Tevell, M. (2000). Driver behavioral adaptation to collision warning and avoidance information. In *Proceedings of the XIVth Triennial Congress of the International Ergonomics Association and 44th Annual Meeting of the Human Factors and Ergonomics Society* (Vol. 3, p. 315). Santa Monica, CA: Human Factors and Ergonomics Society.

de Waard, D. (1996). *The measurement of drivers` mental workload*. (Ph.D. Thesis), University of Groningen, Traffic Research Centre.

de Waard, D., & Brookhuis, K. A. (1991). Assessing driver status: A demonstration experiment on the road. *Accident Analysis and Prevention, 23,* 297–307.

de Waard, D., & Brookhuis, K. A. (1997). On the measurement of driver workload. In T. Rothengatter & E. C. Vaya (Eds.), *Traffic & Transportation Psychology. Theory and Application (pp. 161-171).* Amsterdam: Pergamon.

Engström, J. (2010). The Tactile Detection Task as a Method for Assessing Drivers` Cognitive Load. In G. L. Rupp (Eds.), *Performance Metrics for Assessing Driver Distraction: The Quest for Improved Road Safety (pp. 90-103).* Warrendale, Pensnsylvania: SAE International

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 safety information systems. *Proceedings of the Third International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design*, Rockport, Maine.

Engström, J., & Mårdh, S. (2007). *SafeTE Final Report.* (Report No. 2007:36) Retrieved from http://publikationswebbutik.vv.se/shopping/ShowItem____2907.aspx

Harms, L., & Patten, C. (2003). Peripheral detection as a measure of driver distraction. A study of memory-based versus system-based navigation in a built-up area. *Transportation Research Part F, 6,* 23–36.

Hart, S. G., Staveland, L. E., Hancock, P. A., & Meshkati, N. (1988). *Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research*. Oxford, England: North-Holland.

Human Machine Interface And the Saftey of Traffic in Europe (HASTE), (2004, August 12). HMI and Safety-Related Driver Performance (Report No. HASTE-D2) Retrieved from http://ec.europa.eu/transport/roadsafety_library/publications/haste_d2_v1-3_small.pdf

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,* 255-275.

Jamson, H. (2000). Driving simulation validity: issues of field of view and resolution. *Proceedings of Driving Simulator Conference (DSC 2000)*. Paris, France.

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

Liu, Y. & Wickens, C. D. (1992). Visual scanning with or without spatial uncertainty and divided and selective attention. *Acta Psychologica, 79*, 131-153.

Liu, C.-C., Doong, J.-L., Lin, C.-Y., & Lai, C.-H. (2009). Responses in a manual tracking and visual detection task. *Perceptual and motor Skills, 109,* 899-910.

Manzey, D. (1998). Psychophysiologie mentaler Beanspruchung. In F. Rösler (Ed.), *Ergebnisse und Anwendung der Psychophysiologie (Enzyklopädie der Psychologie). Themenbereich C, Theorie und Forschung. Ergebnisse und Anwendungen der Psychophysiologie* (Vol. 5, pp. 799–864). Göttingen: Hogrefe.

Marshall, S. P., Davis, C., & Knust, S. (2004). *The Index of cognitive Activity: estimating cognitive effort from pupil dilation.* San Diego, CA: Eyetracking Inc.

Martens, M. H., & van Winsum, W. (2000). *Measuring distract ion: The Peripheral Detection Task.* Soesterberg, The Netherlands: TNO Human Factors.

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.

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.

Miura, T. (1986). Coping with situational demands: A study of eyemovements and peripheral vision performance. In A. G. Gale, M. Freeman, C. M. Haslegrave, P. Smith, & S. Taylor, (Eds.), *Vision in Vehicles II.* Amsterdam: Elsevier, North Holland.

Mourant, R. R. & Rockwell, T. H. (1972). Strategies of visual search by novice and experienced drivers. *Human Factors, 14***,** 325-335.

National Highway Traffic Safety Association (NHTSA), Department of Transportation (2012, February 24) Visual-Manual NHTSA Driver Distraction Guidelines for In-Vehicle Electronic Devices (Report No. NHTSA-2010-0053) Retrieved from https://www.federalregister.gov/articles/2012/02/24/2012-4017/visual-manual-nhtsa-driver-distraction-guidelines-for-in-vehicle-electronic-devices

O'Donnell, R. D., & Eggemeier, F. T. (1986). Workload assessment methodology. In K. R. Boff, L. Kaufman, & J. P. Thomas, (Eds.), *Cognitive processes and performance* (Vol. 2, pp. 1-49). Oxford, England: John Wiley & Sons.

Olsson, S., & Burns, P. C. (2000). *Measuring driver visual distraction with a peripheral detection task.* Linköping, Sweden: Linköping University.

Parasuraman, R., & Hancock, P. A. (2001). Adaptive control of mental workload. In P. A. Hancock & P. A. Desmond (Eds.), *Stress, workload, and fatigue*. Mahwah: L. Erlbaum.

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.

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

Rößger, P. (1996). *Die Entwicklung der Pupillometrie zu einer Methode der Messung mentaler Beanspruchung in der Arbeitswissenschaft*. (Dissertation). Sinzheim: Pro Universitate Verlag.

Santos, J., Merat, N., Mouta, S., Brookhuis, K., de Waard, D. (2005). The interaction between driving and in-vehicle information systems: Comparison of results from laboratory, simulator and real-world studies. *Transportation Research Part F, 8*, 135–146.

Schwalm, M., Keinath, A., & Zimmer, H. D. (2008). Pupillometry as a method for measuring mental workload within a simulated driving task. In D. de Waard, F. O. Flemisch, B. Lorenz, H. Oberheid, & K. A. Brookhuis (Eds.), *Human Factors for assistance and automation* (pp. 1-13). Maastricht, the Netherlands: Shaker Publishing.

Schwalm, M. (2009). *Pupillometrie als Methode zur Erfassung mentaler Beanspruchungen im automotiven Kontext*. (Dissertation).Saarbrücken: Univ. Saarbrücken.

Törnros J, & Bolling A. (2006). Mobile phone use – effects of conversation on mental workload and driving speed in rural and urban environments. *Transportation Research Part F, 9,* 298–306.

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 (Eds.), *Performance Metrics for Assessing Driver Distraction: The Quest for Improved Road Safety* (pp. 73-89). Warrendale, Pensnsylvania: SAE International

van Erp, J. B. F., & van Veen, H. A. H. C. (2004). Vibrotactile in-vehicle navigation system. *Transportation Research Part F, 7*, 247–256.

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

Victor, T., Engström, J., & Harbluk, J. (2008). Distraction Assessment Methods Based on Visual Behaviour and Event Detection. In M. A. Regan, J. D. Lee, & K. L. Young, (Eds) *Driver Distraction: Theory, Effects and Mitigation* (pp. 135-165). Boca Raton, Florida: CRC Press.

Wickens, C. D. (1984). Processing resources in attention. In R. Parasuraman, & D. R. Davies (Eds.). *Varieties of attention* (pp. 63–102). San Diego, CA: Academic Press.

Wickens, C. D. (2000). Attention, Time-Sharing, and Workload. In C. D. Wickens & J. G. Hollands (Eds.), *Engineering Psychology and Human performance* (pp. 439-479). Upper Saddle River, NJ: Prentice-Hall.

Wickens, C. D. (2002). Multiple Resources and Performance Prediction. *Theoretical Issues in Ergonomic Science, 3* (2), 159-177.

Wickens, C. D., & Hollands, G. (2000). *Engineering Psychology and Human Performance* (3rd ed.). New Jersey: Prentice Hall.

Wickens, C. D. & Liu, Y. (1988). Codes and modalities in multiple resources: a success and a qualification. *Human Factors, 30*, 599-616.

Wierwille, W. W., & Tijerina, L. (1998). Modeling the relationship between driver in-vehicle visual demands and accident occurrence. In A. G. Gale & I. D. Brown & C. M. Haslegrave & S. Taylor (Eds.), *Vision in Vehicles VI* (pp. 233 - 243). Amsterdam: North-Holland.

Williams, L. J. (1985). Tunnel vision induced by a foveal load manipulation. *Human Factors, 27,* 221–227.

Williams, L. J. (1988). Tunnel vision or general interference? Cognitive load and attentional bias are both important. *American Journal of Psychology, 101,* 171–191.

Zijlstra, F. R. (1993). *Efficiency in work behaviour. A design approach for moderntools.* Delft: Delft University of Technology.

8. Summary in German

In den letzten Jahren wurden vermehrt Fahrerinformations- und Entertainmentsysteme ins Fahrzeug implementiert. Diese dienen dazu, dem Fahrer wertvolle Informationen zu liefern und die Fahrt zu erleichtern. Jedoch stellt sich die Frage, wie sehr der Fahrer durch die Systembedienung tatsächlich abgelenkt wird. Um eine sichere Bedienung während der Fahrt gewährleisten zu können, gilt es den Grad der damit verbundenen Ablenkung zu erfassen. Eine Überbeanspruchung könnte dazu führen, dass fahrrelevante Entscheidungen nicht mehr schnell und sicher getroffen werden und sich damit die Wahrscheinlichkeit für Fahrfehler und Unfälle erhöht.

Gerade während der Entwicklung dieser Systeme ist es deshalb wichtig, dass getestet werden kann, ob und wie sehr diese Systeme gegebenenfalls ablenkend wirken, um dementsprechend die auf den Fahrer einwirkende Belastung so gering wie möglich zu halten.

Die meisten Systeme erfordern eine visuell-manuelle Bedienung und in den vergangenen Jahren wurden eine Reihe von Bewertungsmethoden entwickelt und zum Teil auch ISO-standardisiert (z.b. Okklusionsmethode, ISO 16673), die in der Lage sind diese Komponente der Beanspruchung zu erfassen. Allerdings werden neuerdings vermehrt rein sprachbasierte Informationssysteme implementiert. Somit wird zwar die direkte Ablenkung durch die visuell-manuelle Bedienung während der Fahrt vermieden, jedoch kann die erhöhte mentale Beanspruchung durchaus ebenfalls ablenkend wirken. Die Bewertung und Messung der mentalen Beanspruchung stellt deshalb eine zentrale Herausforderung dar.

Die ursprünglich von Martens und Van Winsum (2000) vorgestellte Peripheral Detection Response Task (PDRT) und die in Folge entwickelten weiteren Varianten der

Detection Response Tasks (DRTs) stellen eine neue, vielversprechende Messmethode dar, um mentale Beanspruchung sensitiv zu erfassen.

In der ursprünglichen Variante, der Peripheral Detection Response Task (PDRT), werden zufällig visuelle Stimuli im peripheren Blickfeld der Versuchsperson präsentiert. Sobald die Versuchsperson solch einen visuellen Reiz entdeckt hat, soll sie dies mittels Druck auf einem Antworttaster signalisieren. Aufgrund der Reaktionszeiten sowie der Detektionsrate ist es möglich Rückschlüsse auf die zugrundeliegende Beanspruchung zu ziehen. Da nicht ausgeschlossen werden kann, dass Kopf- und Augenbewegungen einen nachteiligen Einfluss auf die Wahrnehmung der visuellen Stimuli haben, wurden in weiterer Folge weitere Varianten, wie z.B. die Auditory Detection Response Task (ADRT) sowie die Tactile Detection Response Task (TDRT) entwickelt (Engström, 2010). Hierbei werden die Reize akustisch beziehungsweise haptisch dargeboten.

Bislang wurden diese drei Varianten erst ein einziges Mal gemeinsam in einer Studie untersucht (siehe Merat & Jamson, 2008) und obwohl die Autoren behaupten, dass alle drei Varianten gleich gut geeignet sind, um mentale Beanspruchung zu messen, ist es fraglich, inwiefern diese Ergebnisse generalisierbar sind. Getestet wurden lediglich eine visuell-manuelle, sowie eine kognitive Nebenaufgabe. Weder die Schwierigkeit noch die Darbietungsart wurden hierbei variiert. Laut Wickens` Multipler Ressourcen Theorie (2002) ist jedoch davon auszugehen, dass es zu Wechselwirkungen und Beeinträchtigungen kommt, wenn zwei Aufgaben dieselben Ressourcen benötigen.

Das Ziel dieser Studie war es daher zu beurteilen, ob die DRTs ein sensitives Maß sind, um mentale Beanspruchung zu erfassen, wenn diese systematisch variiert wird und ob es zu Wechselwirkungen zwischen der Darbietungsart der Nebenaufgabe als auch der DRT Variante kommt.

Folgende DRT Varianten wurden in der vorliegenden Studie evaluiert: die Peripheral Detection Response Task (PDRT), die Auditory Detection Response Task (ADRT), sowie die Tactile Detection Response Task (TDRT). Als Nebenaufgabe wurden zwei artifizielle kognitiv beanspruchende Aufgaben gewählt, die in zwei Schwierigkeitsstufen (leicht/schwer) vorgegeben wurden. Zusätzlich wurden diese Nebenaufgaben in drei verschiedenen Darbietungsarten (visuell/auditiv/rein kognitiv) präsentiert, um eben Rückschlüsse darüber zuzulassen, ob es zu Wechselwirkungen zwischen der Darbietungsart und der DRT-Variante kommt und anschließend eine Empfehlung aussprechen zu können.

Insgesamt wurden 24 BMW Mitarbeiter (12 Männer, 12 Frauen) im Forschungs- und Innovationszentrum in München getestet. Der Altersdurchschnitt betrug 29 Jahre mit einer Spannweite von 21 bis 42 Jahren.

Für jede DRT wurde ein 3x2 within subjects Design gewählt, was bedeutet, dass alle Versuchspersonen pro DRT Variante die kognitiven Nebenaufgaben in allen drei Darbietungsformen (visuell, auditiv sowie rein kognitiv), sowie in beiden Schwierigkeitsstufen (leicht, schwer) bearbeitet haben.

Die Ergebnisse bezüglich der Reaktionszeiten zeigten, dass alle DRT Varianten ein sensitives Maß waren, um zwischen niedriger und hoher Beanspruchung klar zu differenzieren. Hierbei ergab sich ein sehr homogenes und kohärentes Bild über alle Darbietungsmodalitäten hinweg. Die Reaktionszeiten in der PDRT waren am kürzesten, gefolgt von denen in der TDT, sowie der ADRT.

Jedoch ergab sich ein anderes Ergebnis, sobald man die Detektionsrate als Bewertungsgrundlage verwendete. Eine signifikante Interaktion zwischen der DRT Variante sowie dem Schwierigkeitsgrad der Nebenaufgabe zeigte, dass bei zunehmender Aufgabenschwierigkeit die DRT Varianten unterschiedlich stark sensitiv reagierten. Eine Detailbetrachtung ergab, dass die PDRT diesbezüglich nicht sensitiv war.

Überraschenderweise waren keinerlei Wechselwirkungen zwischen den DRT Varianten und der Darbietungsart der Nebenaufgabe feststellbar. Dieses Ergebnis steht im Gegensatz zu Wickens` postulierter Multiplen Ressourcen Theorie. Eine mögliche Erklärung hierfür ist, dass die DRTs selbst so wenig Beanspruchung erfordern, dass es zu keiner Beeinträchtigung kam. Auch könnten die Nebenaufgaben, selbst in der schwierigen Ausprägung, zu einfach gewesen sein, so dass die Versuchspersonen nicht ihr Limit bei der Informationsverarbeitung erreicht haben.

Gemäß den Ergebnissen dieser Studie wird der Gebrauch der TDRT sowie der ADRT empfohlen, da beide sowohl in Bezug auf die Reaktionszeiten als auch die Detektionsrate ein sensitives Maß waren, um mentale Beanspruchung zu messen.

Jedoch sind weitere Studien notwendig, die auch Realaufgaben untersuchen, um die DRTs letztendlich als valide Messmethode für mentale Beanspruchung zu etablieren.

9. Appendix

9.1. Materials

9.1.1. Counting Task – Paper cards

Anleitung Zählen_leicht_1

Bitte zählen Sie in **2er** Schritten **aufwärts**, ausgehend von dieser Zahl:

463

Bitte zählen Sie in **2er** Schritten **aufwärts**, ausgehend von dieser Zahl:

544

Bitte zählen Sie in **2er** Schritten **aufwärts**, ausgehend von dieser Zahl:

623

Anleitung Zählen_leicht_2

Bitte zählen Sie in **2er** Schritten **aufwärts**, ausgehend von dieser Zahl:

175

Bitte zählen Sie in **2er** Schritten **aufwärts**, ausgehend von dieser Zahl:

256

Bitte zählen Sie in **2er** Schritten **aufwärts**, ausgehend von dieser Zahl:

335

Anleitung Zählen_leicht_3

Bitte zählen Sie in **2er** Schritten **aufwärts**, ausgehend von dieser Zahl:

254

Bitte zählen Sie in **2er** Schritten **aufwärts**, ausgehend von dieser Zahl:

335

Bitte zählen Sie in **2er** Schritten **aufwärts**, ausgehend von dieser Zahl:

414

Anleitung Zählen_schwer_1

Bitte zählen Sie in **7er** Schritten **abwärts**, ausgehend von dieser Zahl:

887

Bitte zählen Sie in **7er** Schritten **abwärts**, ausgehend von dieser Zahl:

606

Bitte zählen Sie in **7er** Schritten **abwärts**, ausgehend von dieser Zahl:

327

Anleitung Zählen_schwer_2

Bitte zählen Sie in **7er** Schritten **abwärts**, ausgehend von dieser Zahl:

976

Bitte zählen Sie in **7er** Schritten **abwärts**, ausgehend von dieser Zahl:

695

Bitte zählen Sie in **7er** Schritten **abwärts**, ausgehend von dieser Zahl:

414

Anleitung Zählen_schwer_3

Bitte zählen Sie in **7er** Schritten **abwärts**, ausgehend von dieser Zahl:

869

Bitte zählen Sie in **7er** Schritten **abwärts**, ausgehend von dieser Zahl:

588

Bitte zählen Sie in **7er** Schritten **abwärts**, ausgehend von dieser Zahl:

307

9.1.2. n-back task – Audio files

Please see the attached CD for the n-back task audio files.

Auditiv_Set_1_leicht

Nächste Folge	8	7	4	5	2	3	1	9	6	0
Nächste Folge	7	3	6	4	0	5	8	1	9	2
Nächste Folge	2	5	3	4	8	0	7	1	9	6

Auditiv_Set_2_leicht

Nächste Folge	6	5	7	0	1	2	9	8	3	4
Nächste Folge	9	2	5	3	7	8	1	6	0	4
Nächste Folge	1	6	7	0	3	9	4	5	2	8

Auditiv_Set_3_leicht

Nächste Folge	7	6	0	2	1	3	5	9	4	8
Nächste Folge	0	4	3	7	5	9	8	1	2	6
Nächste Folge	3	5	8	1	9	6	0	4	2	7

Auditiv_Set_1_schwer

Nächste Folge	6	5	2	3	8	1	4	9	0	7
Nächste Folge	5	7	0	8	3	2	6	4	1	9
Nächste Folge	9	6	0	2	3	5	8	1	7	4

Auditiv_Set_2_schwer

Nächste Folge	6	0	3	8	5	9	7	1	2	4
Nächste Folge	4	2	0	3	9	6	5	1	7	8
Nächste Folge	8	5	6	7	9	2	3	4	0	1

Auditiv_Set_3_schwer

Nächste Folge	4	7	0	9	5	3	6	2	1	8
Nächste Folge	9	0	1	7	3	2	6	8	4	5
Nächste Folge	9	5	1	7	8	3	4	6	0	2

9.1.3. n-back task – PowerPoint presentations

Please see the attached CD for the n-back task PowerPoint presentations.

Visuell_Set_1_leicht

Nächste Folge	9	2	0	8	7	5	3	1	6	4	Ende
Nächste Folge	0	5	9	1	3	8	7	6	4	2	Ende
Nächste Folge	5	2	8	4	1	3	6	0	7	9	Ende

Visuell_Set_2_leicht

Nächste Folge	2	4	8	3	5	6	9	7	1	0	Ende
Nächste Folge	0	9	4	6	8	1	3	2	5	7	Ende
Nächste Folge	2	4	7	9	1	8	0	6	5	3	Ende

Visuell_Set_3_leicht

Nächste Folge	7	2	6	0	3	8	5	4	1	9	Ende
Nächste Folge	4	9	0	3	6	1	7	5	2	8	Ende
Nächste Folge	3	0	4	5	6	2	9	7	1	8	Ende

Visuell_Set_1_schwer

Nächste Folge	6	9	7	8	3	2	4	1	0	5	Ende
Nächste Folge	2	5	6	8	1	0	3	4	7	9	Ende
Nächste Folge	0	6	7	4	2	1	3	9	8	5	Ende

Visuell_Set_2_schwer

Nächste Folge	5	7	0	4	2	1	3	9	8	6	Ende
Nächste Folge	0	3	2	7	1	4	5	9	6	8	Ende
Nächste Folge	1	2	3	5	8	0	6	4	9	7	Ende

Visuell_Set_3_schwer

Nächste Folge	3	8	4	5	1	7	0	2	9	6	Ende
Nächste Folge	0	5	7	6	1	8	9	3	2	4	Ende
Nächste Folge	2	9	8	4	0	3	7	5	6	1	Ende

9.2. Statistical Analysis

9.2.1. Extract from data sheet

	Geschlecht	Alter	Händigkeit	Taster	Vibrator	Reaktionszeit PDT Varianten	PDRT_V_L_1 RT	PDRT_V_L_2 RT	PDRT_V_L_3 RT	**PDRT V L**	gesamt
VP 1	m	33	rechts	rechts	links		662	561,6666667	446,6666667	**504,1666667**	556,7777778
VP 2	w	25	rechts	rechts	links		476,6666667	484	425	**454,5**	461,8888889
VP 3	w	23	rechts	rechts	links		602	566	613,3333333	**589,6666667**	593,7777778
VP 4	w	26	rechts	rechts	links		755	798,3333333	733,3333333	**765,8333333**	762,2222222
VP 5	m	33	rechts	rechts	links		452	551,6666667	428,3333333	**490**	477,3333333
VP 6	m	32	links	links	rechts		468,3333333	438,3333333	522	**480,1666667**	476,2222222
VP 7	m	27	rechts	rechts	links		408	423,3333333	484	**453,6666667**	438,4444444
VP 8	w	29	rechts	rechts	links		431,6666667	493,3333333	404	**448,6666667**	443
VP 9	w	23	rechts	rechts	links		512	542	430	**486**	494,6666667
VP 10	w	31	rechts	rechts	links		546,6666667	566,6666667	556,6666667	**561,6666667**	556,6666667
VP 11	w	26	rechts	rechts	links		646	611,6666667	741,6666667	**676,6666667**	666,4444444
VP 12	m	26	rechts	rechts	links		754	460	454	**457**	556
VP 13	m	32	rechts	rechts	links		738	418,3333333	483,3333333	**450,8333333**	546,5555556
VP 14	m	28	rechts	rechts	links		458,3333333	468,3333333	330	**399,1666667**	418,8888889
VP 15	w	26	links	links	rechts		548	470	382	**426**	466,6666667
VP 16	w	26	rechts	rechts	links		478,3333333	398	616,6666667	**507,3333333**	497,6666667
VP 17	m	32	rechts	rechts	links		358	321,6666667	355	**338,3333333**	344,8888889
VP 18	w	32	rechts	rechts	links		601,6666667	610	590	**600**	600,5555556
VP 19	w	42	rechts	rechts	links		440	435	448	**441,5**	441
VP 20	w	25	rechts	rechts	links		372	346,6666667	311,6666667	**329,1666667**	343,4444444
VP 21	m	36	links	rechts	links		713,3333333	708	828,3333333	**768,1666667**	749,8888889
VP 22	m	21	rechts	rechts	links		426,6666667	496,6666667	544	**520,3333333**	489,1111111
VP 23	m	24	rechts	rechts	links		366,6666667	330	355	**342,5**	350,5555556
VP 24	m	38	rechts	rechts	links		332	555	498	**526,5**	461,6666667

9.2.2. Description of the sample

Table 5. Age distribution.

	N	Minimum	Maximum	Mean	Std. Deviation
Alter	24	21	42	29,00	5,167
Valid N (listwise)	24				

Table 6. Gender distribution.

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	m	12	50,0	50,0	50,0
	w	12	50,0	50,0	100,0
	Total	24	100,0	100,0	

Table 7. Handedness distribution.

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	links	3	12,5	12,5	12,5
	rechts	21	87,5	87,5	100,0
	Total	24	100,0	100,0	

9.2.3. Reaction Time – visual presentation mode

Table 8. Descriptive Statistics of the Reaction Times of the different DRT variants in the visual presentation mode.

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
PDRT V L	24	329,16667	768,16667	500,7431	115,68476009
PDRT V S	24	386,16667	1611,250	618,2341	240,35891534
TDRT V L	24	397,00000	1155,500	607,8264	181,26546897
TDRT V S	24	421,83333	1224,000	754,1701	207,77962429
ADRT V L	24	488,83333	1157,500	741,3413	156,70578933
ADRT V S	24	503,16667	1343,250	839,9365	215,97204856
Valid N (listwise)	24				

Table 9. General Linear Model – Reaction Time.

Tests der Innersubjekteffekte

Maß:MASS_1

Quelle		Quadratsumme vom Typ III	df	Mittel der Quadrate	F	Sig.
ArtDRT	Sphärizität angenommen	1283458,036	2	641729,018	41,945	,000
	Greenhouse-Geisser	1283458,036	1,771	724802,606	41,945	,000
	Huynh-Feldt	1283458,036	1,908	672786,783	41,945	,000
	Untergrenze	1283458,036	1,000	1283458,036	41,945	,000
Fehler(ArtDRT)	Sphärizität angenommen	703760,217	46	15299,135		
	Greenhouse-Geisser	703760,217	40,728	17279,650		
	Huynh-Feldt	703760,217	43,877	16039,568		
	Untergrenze	703760,217	23,000	30598,270		
Schwierigkeitgrad	Sphärizität angenommen	525422,192	1	525422,192	32,675	,000
	Greenhouse-Geisser	525422,192	1,000	525422,192	32,675	,000
	Huynh-Feldt	525422,192	1,000	525422,192	32,675	,000
	Untergrenze	525422,192	1,000	525422,192	32,675	,000
Fehler(Schwierigkeitgrad)	Sphärizität angenommen	369850,391	23	16080,452		
	Greenhouse-Geisser	369850,391	23,000	16080,452		
	Huynh-Feldt	369850,391	23,000	16080,452		
	Untergrenze	369850,391	23,000	16080,452		
ArtDRT * Schwierigkeitgrad	Sphärizität angenommen	13877,800	2	6938,900	,624	,540
	Greenhouse-Geisser	13877,800	1,836	7559,532	,624	,527
	Huynh-Feldt	13877,800	1,987	6983,306	,624	,539
	Untergrenze	13877,800	1,000	13877,800	,624	,438
Fehler(Art DRT*Schwierigkeitgrad)	Sphärizität angenommen	511382,992	46	11117,022		
	Greenhouse-Geisser	511382,992	42,223	12111,355		
	Huynh-Feldt	511382,992	45,707	11188,166		
	Untergrenze	511382,992	23,000	22234,043		

Table 10. T-Test Reaction Time of all DRT variants under the visual presentation mode.

Paired Samples Test

		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference Lower	Upper			
Pair 1	PDRT V L - PDRT V S	-117,491	182,87147870	37,32848	-194,711	-40,2712	-3,147	23	,005
Pair 2	TDRT V L - TDRT V S	-146,344	151,86627316	30,99957	-210,471	-82,2162	-4,721	23	,000
Pair 3	ADRT V L - ADRT V S	-98,5952	141,85784163	28,95661	-158,497	-38,6939	-3,405	23	,002

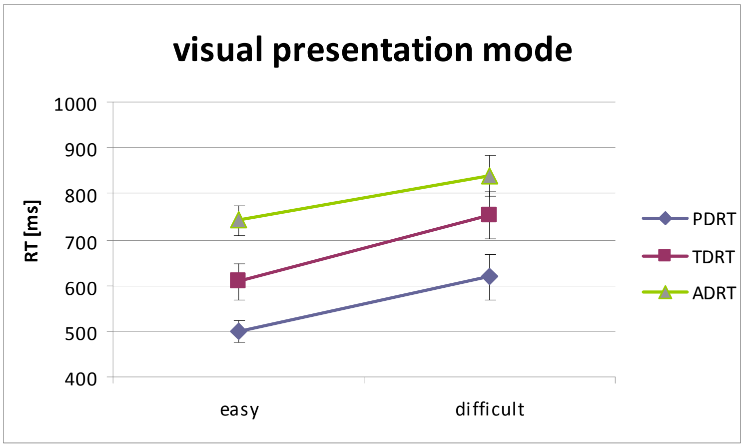

Figure 11. Compared Reaction Times of the different DRT variants under the visual presentation mode.

9.2.4. Reaction Time – auditory presentation mode

Table 11. Descriptive Statistics of the Reaction Times of the different DRT variants in the auditory presentation mode.

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
PDRT A L	24	313,66667	697,00000	485,5938	114,24828171
PDRT A S	24	343,00000	938,00000	599,8889	170,72674064
TDRT A L	24	381,00000	774,25000	549,0799	106,52062432
TDRT A S	24	437,00000	1060,000	711,9306	163,99425705
ADRT A L	24	497,00000	1163,000	710,8889	165,74637890
ADRT A S	24	589,00000	1235,000	870,3576	182,50006862
Valid N (listwise)	24				

Table 12. General Linear Model – Reaction Time.

Tests der Innersubjekteffekte

Maß:MASS_1

Quelle		Quadratsumme vom Typ III	df	Mittel der Quadrate	F	Sig.
ArtDRT	Sphärizität angenommen	1516572,005	2	758286,002	92,692	,000
	Greenhouse-Geisser	1516572,005	1,958	774387,350	92,692	,000
	Huynh-Feldt	1516572,005	2,000	758286,002	92,692	,000
	Untergrenze	1516572,005	1,000	1516572,005	92,692	,000
Fehler(ArtDRT)	Sphärizität angenommen	376314,331	46	8180,746		
	Greenhouse-Geisser	376314,331	45,044	8354,455		
	Huynh-Feldt	376314,331	46,000	8180,746		
	Untergrenze	376314,331	23,000	16361,493		
Schwierigkeitgrad	Sphärizität angenommen	762529,178	1	762529,178	129,445	,000
	Greenhouse-Geisser	762529,178	1,000	762529,178	129,445	,000
	Huynh-Feldt	762529,178	1,000	762529,178	129,445	,000
	Untergrenze	762529,178	1,000	762529,178	129,445	,000
Fehler(Schwierigkeitgrad)	Sphärizität angenommen	135487,490	23	5890,760		
	Greenhouse-Geisser	135487,490	23,000	5890,760		
	Huynh-Feldt	135487,490	23,000	5890,760		
	Untergrenze	135487,490	23,000	5890,760		
ArtDRT * Schwierigkeitgrad	Sphärizität angenommen	17638,939	2	8819,469	2,149	,128
	Greenhouse-Geisser	17638,939	1,774	9941,066	2,149	,135
	Huynh-Feldt	17638,939	1,912	9225,164	2,149	,131
	Untergrenze	17638,939	1,000	17638,939	2,149	,156
Fehler(Art DRT*Schwierigkeitgrad)	Sphärizität angenommen	188746,578	46	4103,186		
	Greenhouse-Geisser	188746,578	40,810	4625,000		
	Huynh-Feldt	188746,578	43,977	4291,933		
	Untergrenze	188746,578	23,000	8206,373		

Table 13. T-Test Reaction Time of all DRT variants under the auditory presentation mode.

Paired Samples Test

		Paired Differences					t	df	Sig. (2-tailed)
					95% Confidence Interval of the Difference				
		Mean	Std. Deviation	Std. Error Mean	Lower	Upper			
Pair 1	PDRT A L - PDRT A S	-114,295	82,28008671	16,79535	-149,039	-79,5513	-6,805	23	,000
Pair 2	TDRT A L - TDRT A S	-162,851	119,66660858	24,42684	-213,381	-112,320	-6,667	23	,000
Pair 3	ADRT A L - ADRT A S	-159,469	84,28616074	17,20484	-195,060	-123,878	-9,269	23	,000

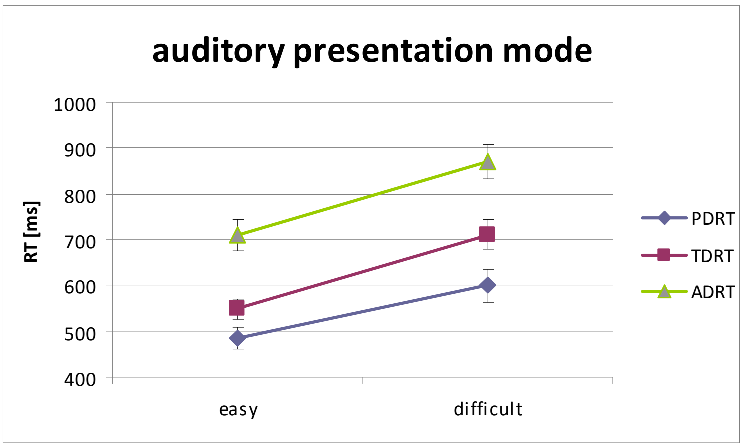

Figure 12. Compared Reaction Times of the different DRT variants under the auditory presentation mode.

9.2.5. Reaction Time – cognitive presentation mode

Table 14. Descriptive Statistics of the Reaction Times of the different DRT variants in the cognitive presentation mode.

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
PDRT C L	24	329,16667	887,42857	508,0248	143,43153685
PDRT C S	24	326,78571	1024,286	589,4578	174,26918703
TDRT C L	24	390,71429	1114,643	599,6910	188,45116541
TDRT C S	24	383,57143	1430,000	721,6741	253,48018821
ADRT C L	24	482,50000	1052,500	705,2227	157,95568462
ADRT C S	24	518,69048	1202,667	776,7349	207,48997429
Valid N (listwise)	24				

Table 15. General Linear Model – Reaction Time.

Tests der Innersubjekteffekte

Maß:MASS_1

Quelle		Quadratsumme vom Typ III	df	Mittel der Quadrate	F	Sig.
ArtDRT	Sphärizität angenommen	894937,156	2	447468,578	39,896	,000
	Greenhouse-Geisser	894937,156	1,838	486910,402	39,896	,000
	Huynh-Feldt	894937,156	1,990	449723,595	39,896	,000
	Untergrenze	894937,156	1,000	894937,156	39,896	,000
Fehler(ArtDRT)	Sphärizität angenommen	515931,898	46	11215,911		
	Greenhouse-Geisser	515931,898	42,274	12204,530		
	Huynh-Feldt	515931,898	45,769	11272,433		
	Untergrenze	515931,898	23,000	22431,822		
Schwierigkeitgrad	Sphärizität angenommen	302342,332	1	302342,332	29,146	,000
	Greenhouse-Geisser	302342,332	1,000	302342,332	29,146	,000
	Huynh-Feldt	302342,332	1,000	302342,332	29,146	,000
	Untergrenze	302342,332	1,000	302342,332	29,146	,000
Fehler(Schwierigkeitgrad)	Sphärizität angenommen	238585,310	23	10373,274		
	Greenhouse-Geisser	238585,310	23,000	10373,274		
	Huynh-Feldt	238585,310	23,000	10373,274		
	Untergrenze	238585,310	23,000	10373,274		
ArtDRT * Schwierigkeitgrad	Sphärizität angenommen	17160,218	2	8580,109	,812	,450
	Greenhouse-Geisser	17160,218	1,609	10666,719	,812	,428
	Huynh-Feldt	17160,218	1,711	10026,640	,812	,434
	Untergrenze	17160,218	1,000	17160,218	,812	,377
Fehler(Art DRT*Schwierigkeitgrad)	Sphärizität angenommen	486161,499	46	10568,728		
	Greenhouse-Geisser	486161,499	37,002	13138,954		
	Huynh-Feldt	486161,499	39,364	12350,523		
	Untergrenze	486161,499	23,000	21137,456		

Table 16. T-Test Reaction Time of all DRT variants under the cognitive presentation mode.

Paired Samples Test

		Paired Differences					t	df	Sig. (2-tailed)
					95% Confidence Interval of the Difference				
		Mean	Std. Deviation	Std. Error Mean	Lower	Upper			
Pair 1	PDRT C L - PDRT C S	-81,4330	122,77049322	25,06042	-133,274	-29,5916	-3,249	23	,004
Pair 2	TDRT C L - TDRT C S	-121,983	173,74345535	35,46523	-195,349	-48,6177	-3,440	23	,002
Pair 3	ADRT C L - ADRT C S	-71,5122	133,27445136	27,20453	-127,789	-15,2353	-2,629	23	,015

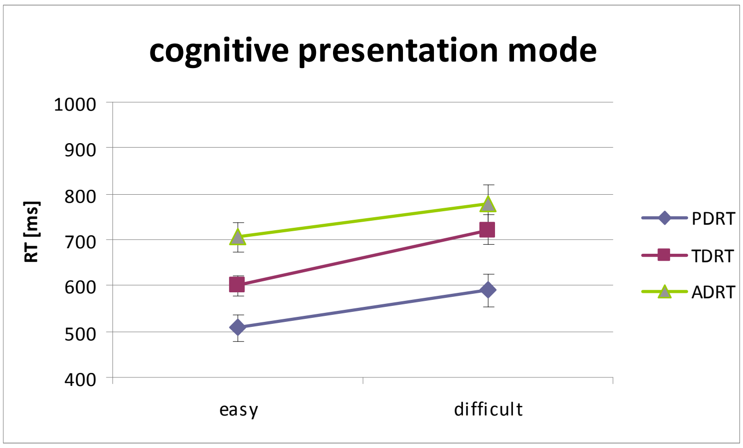

Figure 13. Compared Reaction Times of the different DRT variants under the cognitive presentation mode.

9.2.6. Hit Rate – visual presentation mode

Table 17. Descriptive Statistics of the Hit Rate of the different DRT variants in the visual presentation mode.

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
PDRT V L	24	92	100	98,61	3,172
PDRT V S	24	37	100	91,98	14,470
TDRT V L	24	90	100	98,94	2,900
TDRT V S	24	42	100	90,58	15,832
ADRT V L	24	73	100	96,46	7,144
ADRT V S	24	55	100	89,01	13,333
Valid N (listwise)	24				

Table 18. General Linear Model – Hit Rate.

Tests der Innersubjekteffekte

Maß:MASS_1

Quelle		Quadratsumme vom Typ III	df	Mittel der Quadrate	F	Sig.
ArtDRT	Sphärizität angenommen	175,435	2	87,718	1,317	,278
	Greenhouse-Geisser	175,435	1,959	89,540	1,317	,278
	Huynh-Feldt	175,435	2,000	87,718	1,317	,278
	Untergrenze	175,435	1,000	175,435	1,317	,263
Fehler(ArtDRT)	Sphärizität angenommen	3064,096	46	66,611		
	Greenhouse-Geisser	3064,096	45,064	67,994		
	Huynh-Feldt	3064,096	46,000	66,611		
	Untergrenze	3064,096	23,000	133,222		
Schwierigkeitsgrad	Sphärizität angenommen	2014,300	1	2014,300	13,475	,001
	Greenhouse-Geisser	2014,300	1,000	2014,300	13,475	,001
	Huynh-Feldt	2014,300	1,000	2014,300	13,475	,001
	Untergrenze	2014,300	1,000	2014,300	13,475	,001
Fehler (Schwierigkeitsgrad)	Sphärizität angenommen	3438,194	23	149,487		
	Greenhouse-Geisser	3438,194	23,000	149,487		
	Huynh-Feldt	3438,194	23,000	149,487		
	Untergrenze	3438,194	23,000	149,487		
ArtDRT * Schwierigkeitsgrad	Sphärizität angenommen	18,100	2	9,050	,161	,852
	Greenhouse-Geisser	18,100	1,942	9,319	,161	,846
	Huynh-Feldt	18,100	2,000	9,050	,161	,852
	Untergrenze	18,100	1,000	18,100	,161	,692
Fehler(Art DRT*Schwierigkeitsgrad)	Sphärizität angenommen	2593,691	46	56,385		
	Greenhouse-Geisser	2593,691	44,672	58,060		
	Huynh-Feldt	2593,691	46,000	56,385		
	Untergrenze	2593,691	23,000	112,769		

Table 19. T-Test Hit Rate of all DRT variants under the visual presentation mode.

Paired Samples Test

		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference Lower	Upper			
Pair 1	PDRT V L - PDRT V S	6,627	14,034	2,865	,701	12,553	2,313	23	,030
Pair 2	TDRT V L - TDRT V S	8,363	14,125	2,883	2,399	14,328	2,901	23	,008
Pair 3	ADRT V L - ADRT V S	7,450	11,315	2,310	2,673	12,228	3,226	23	,004

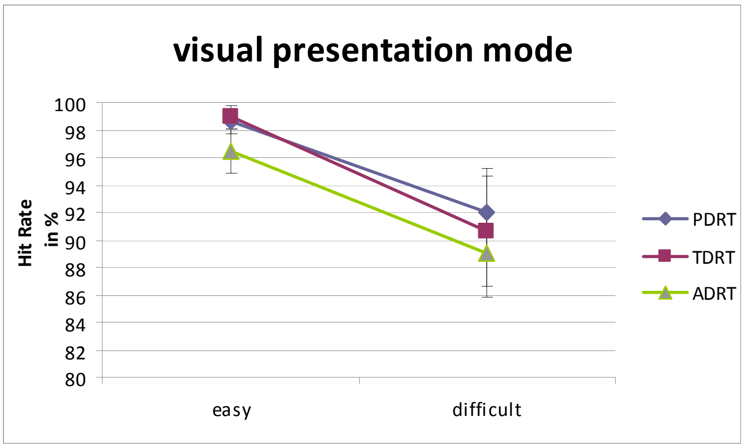

Figure 14. Compared Hit Rates of the different DRT variants under the visual presentation mode.

9.2.7. Hit Rate – auditory presentation mode

Table 20. Descriptive Statistics of the Hit Rates of the different DRT variants in the auditory presentation mode.

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
PDRT A L	24	90	100	98,89	3,016
PDRT A S	24	73	100	92,64	8,511
TDRT A L	24	90	100	99,24	2,600
TDRT A S	24	55	100	90,69	11,239
ADRT A L	24	83	100	96,74	4,975
ADRT A S	24	47	100	88,68	15,510
Valid N (listwise)	24				

Table 21. General Linear Model – Hit Rate.

Tests der Innersubjekteffekte

Maß:MASS_1

Quelle		Quadratsumme vom Typ III	df	Mittel der Quadrate	F	Sig.
ArtDRT	Sphärizität angenommen	241,088	2	120,544	2,591	,086
	Greenhouse-Geisser	241,088	1,983	121,595	2,591	,086
	Huynh-Feldt	241,088	2,000	120,544	2,591	,086
	Untergrenze	241,088	1,000	241,088	2,591	,121
Fehler(ArtDRT)	Sphärizität angenommen	2140,394	46	46,530		
	Greenhouse-Geisser	2140,394	45,602	46,936		
	Huynh-Feldt	2140,394	46,000	46,530		
	Untergrenze	2140,394	23,000	93,061		
Schwierigkeitsgrad	Sphärizität angenommen	2087,982	1	2087,982	19,554	,000
	Greenhouse-Geisser	2087,982	1,000	2087,982	19,554	,000
	Huynh-Feldt	2087,982	1,000	2087,982	19,554	,000
	Untergrenze	2087,982	1,000	2087,982	19,554	,000
Fehler (Schwierigkeitsgrad)	Sphärizität angenommen	2455,999	23	106,783		
	Greenhouse-Geisser	2455,999	23,000	106,783		
	Huynh-Feldt	2455,999	23,000	106,783		
	Untergrenze	2455,999	23,000	106,783		
ArtDRT * Schwierigkeitsgrad	Sphärizität angenommen	34,992	2	17,496	,302	,741
	Greenhouse-Geisser	34,992	1,804	19,392	,302	,719
	Huynh-Feldt	34,992	1,949	17,955	,302	,735
	Untergrenze	34,992	1,000	34,992	,302	,588
Fehler(Art DRT*Schwierigkeitsgrad)	Sphärizität angenommen	2666,860	46	57,975		
	Greenhouse-Geisser	2666,860	41,503	64,257		
	Huynh-Feldt	2666,860	44,824	59,496		
	Untergrenze	2666,860	23,000	115,950		

Table 22. T-Test Hit Rate of all DRT variants under the auditory presentation mode.

Paired Samples Test

		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference Lower	Upper			
Pair 1	PDRT A L - PDRT A S	6,250	9,158	1,869	2,383	10,117	3,343	23	,003
Pair 2	TDRT A L - TDRT A S	8,542	10,136	2,069	4,262	12,822	4,129	23	,000
Pair 3	ADRT A L - ADRT A S	8,056	16,089	3,284	1,262	14,849	2,453	23	,022

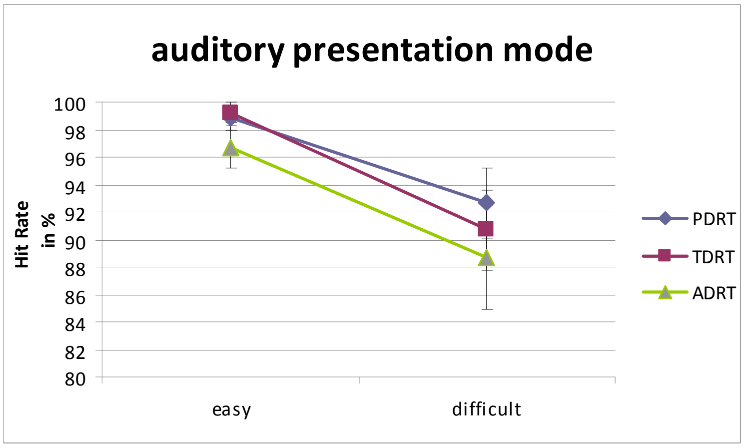

Figure 15. Compared Hit Rates of the different DRT variants under the auditory presentation mode.

9.2.8. Hit Rate – cognitive presentation mode

Table 23. Descriptive Statistics of the Hit Rates of the different DRT variants in the cognitive presentation mode.

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
PDRT C L	24	92	100	98,07	3,436
PDRT C S	24	83	100	95,58	6,097
TDRT C L	24	75	100	94,76	7,746
TDRT C S	24	48	100	85,17	16,070
ADRT C L	24	90	100	96,40	4,058
ADRT C S	24	67	100	89,47	10,340
Valid N (listwise)	24				

Table 24. General Linear Model – Hit Rate.

Tests der Innersubjekteffekte

Maß:MASS_1

Quelle		Quadratsumme vom Typ III	df	Mittel der Quadrate	F	Sig.
ArtDRT	Sphärizität angenommen	1134,838	2	567,419	6,428	,003
	Greenhouse-Geisser	1134,838	1,399	810,897	6,428	,009
	Huynh-Feldt	1134,838	1,462	776,034	6,428	,008
	Untergrenze	1134,838	1,000	1134,838	6,428	,018
Fehler(ArtDRT)	Sphärizität angenommen	4060,471	46	88,271		
	Greenhouse-Geisser	4060,471	32,188	126,148		
	Huynh-Feldt	4060,471	33,634	120,725		
	Untergrenze	4060,471	23,000	176,542		
Schwierigkeitsgrad	Sphärizität angenommen	1444,829	1	1444,829	18,437	,000
	Greenhouse-Geisser	1444,829	1,000	1444,829	18,437	,000
	Huynh-Feldt	1444,829	1,000	1444,829	18,437	,000
	Untergrenze	1444,829	1,000	1444,829	18,437	,000
Fehler (Schwierigkeitsgrad)	Sphärizität angenommen	1802,455	23	78,368		
	Greenhouse-Geisser	1802,455	23,000	78,368		
	Huynh-Feldt	1802,455	23,000	78,368		
	Untergrenze	1802,455	23,000	78,368		
ArtDRT * Schwierigkeitsgrad	Sphärizität angenommen	308,774	2	154,387	3,383	,043
	Greenhouse-Geisser	308,774	1,670	184,862	3,383	,052
	Huynh-Feldt	308,774	1,786	172,921	3,383	,049
	Untergrenze	308,774	1,000	308,774	3,383	,079
Fehler(Art DRT*Schwierigkeitsgrad)	Sphärizität angenommen	2099,439	46	45,640		
	Greenhouse-Geisser	2099,439	38,417	54,649		
	Huynh-Feldt	2099,439	41,070	51,119		
	Untergrenze	2099,439	23,000	91,280		

Table 25. T-Test Hit Rate of all DRT variants under the cognitive presentation mode.

Paired Samples Test

		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference Lower	Upper			
Pair 1	PDRT C L - PDRT C S	2,488	6,356	1,297	-,196	5,172	1,917	23	,068
Pair 2	TDRT C L - TDRT C S	9,587	13,600	2,776	3,844	15,329	3,453	23	,002
Pair 3	ADRT C L - ADRT C S	6,931	10,674	2,179	2,424	11,439	3,181	23	,004

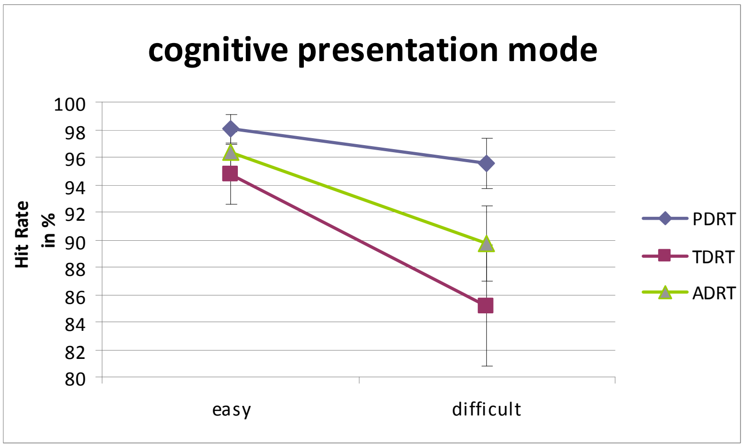

Figure 16. Compared Hit Rates of the different DRT variants under the visual presentation mode.

9.3. Curriculum Vitae

Lebenslauf
Juliane Schäfer

PERSÖNLICHE DATEN

Adresse	Schottenfeldgasse 66/1/18 1070 Wien Österreich
Telefon	+43 650 7158833
e-mail	julianeschaefer@gmx.de
Geburtsdatum, -ort	28.10.1985, Dachau
Staatsangehörigkeit	deutsch
Familienstand	ledig

STUDIUM/AUSBILDUNG

Studiengang Psychologie	angestrebter Abschluss: Diplom-Psychologe (voraussichtlich Juni 2012)
10.2005 – dato	Universität Wien
Schulausbildung	Abschluss: Allgemeine Hochschulreife (2,3), 24.06.2005
09.2003 – 06.2005	Werner-Heisenberg Gymnasium, Garching bei München
09.1997 – 07.2003	Carl-Orff-Gymnasium, Unterschleißheim

PRAKTIKA/BERUFSERFAHRUNG

10.2009 – dato	Firma Empirience, Regensburg Subunternehmer, Social Media Analyst für VS-25 Internetrecherche, Transkription und Aufbereitung von Nutzermeinungen, Erstellung einer wöchentlichen Zusammenfassung
08.2009 – 09.2009	BMW Group, München Abteilung „Anzeige- und Bedienkonzepte“; (EI-231) Analyse und Aufbereitung von Kundenäußerungen, Abstimmung der Erhebungsinstrumente zur Nutzerbefragung
11.2008 – 06.2009	BMW Technology Office, Palo Alto, Kalifornien Abteilung ZT-Z MINI E Kundenstudie: Mithilfe bei der Erstellung der Erhebungsinstrumente, Koordination zwischen deutschen und amerikanischen Partnern und Kollegen, Social Media Analyse der Kundenmeinungen

08.2009 – 09.2009	BMW Group, München Abteilung „Anzeige- und Bedienkonzepte“; (EI-231) Analyse und Aufbereitung von Kundenäußerungen, Abstimmung der Erhebungsinstrumente zur Nutzerbefragung
07.2008 – 09.2008	BMW Group, München Abteilung „Anzeige- und Bedienkonzepte“; (EI-23) Versuchsplanung, -durchführung und –auswertung von zwei Probandenversuchen im Usability Labor und Fahrsimulator
07.2007 – 08.2007	Firma Empirience, Regensburg Mitwirkung Versuchsplanung, Versuchsdurchführung und Auswertung eines Feldversuchs zum Einpark-Assistenten

DIPLOMARBEIT

07.2011 – 12.2011	BMW Group, München Abteilung „Anzeige- und Bedienkonzepte“; (EI-211) „Detection Response Tasks – A New Method to Measure Cognitive Workload“ Studienkonzeption, technischer Aufbau, Versuchsdurchführung, quantitative Datenanalyse (SPSS)

KENNTNISSE UND FÄHIGKEITEN

Computerkenntnisse	ECDL Core (Europäischer Computerführerschein), MS Office (sehr gut), SPSS (sehr gut), PHP (Grundkenntnisse), Java (Grundkenntnisse), 10 Finger Maschinenschreiben (sehr gut),
Weitere Kenntnisse	Coaching-Seminar, Problemorientiertes Seminar „Entwicklung einer Weiterbildungsmaßnahme zur Umsetzung für die Praxis“

FREMDSPRACHEN

Fremdsprachen	Englisch (verhandlungssicher), Französisch (Grundkenntnisse), Spanisch (Grundkenntnisse

Wien, 09.05.2012