Effects of Task and Presentation Modality in Detection Response Tasks

Roman Vilimek, Juliane Schäfer, and Andreas Keinath

BMW Group, Concept Quality, Munich, Germany
{roman.vilimek,juliane.schaefer,andreas.keinath}@bmw.de

Abstract. To assess driver distraction adequately, cognitive workload measurement techniques are necessary that can be used as part of standard in-vehicle testing procedures. Detection response tasks (DRTs) are a simple and effective way of assessing workload. However, as DRTs require cognitive resources themselves, interferences between task modality and DRT modality are possible. In this study, DRT stimuli (auditory, visual, tactile) are varied systematically with secondary task presentation modality (auditory, visual, or purely cognitive tasks). The aim is to infer if different DRT variants remain sensitive to changes in workload even if primary and secondary task convey information using the same presentation modality, thus making resource conflicts likely. Results show that all DRTs successfully discriminate between high and low workload levels in terms of reaction time independent of DRT presentation modality. Differences in discriminability can be found in hit rate measurement.

Keywords: DRT, PDT, workload, driver distraction.

1 Introduction

Driver distraction is in most cases defined as inadequate visual-manual orientation to the driving task. Usability tests in the automotive context typically use eye tracking and defined acceptable thresholds for eyes-off-the-road durations (e.g. [1]). However, in attempts to reduce driver distraction, recent developments in in-vehicle information and communication systems show a trend towards less visual-manual activity. This trend includes technologies like speech interaction that reduces visual-manual demands but may increase cognitive demands. In order to assess the cognitive distraction potential of these new technologies in a suitable way during standard usability testing procedures, a simple continuous method is required which is able to reliably differentiate between different levels of cognitive task demands.

Self-reported measures of workload are based upon the premise that a test participant is able to identify and state the experienced load of a task at hand. Subjective rating scales like the unidimensional Rating Scale of Mental Effort (RSME, [2]) or the multidimensional NASA Task Load Index (TLX, [3]) are in most cases administered directly after task processing and proved to be sensitive to changes in workload. A disadvantage of these methods is their summative and a posteriori

D. Harris (Ed.): EPCE/HCII 2013, Part I, LNAI 8019, pp. 177–185, 2013.

nature. Subjects have to recall at the time of the survey the degree of cognitive workload they experienced during completing a certain task. If tasks are complex like using technical systems, short peaks in cognitive effort can be over- or underrated or even completely forgotten. Therefore, if the objective of a usability test is to evaluate specific aspects of a complex technical system, single use cases need to be isolated to be assessed individually which often makes the procedure more difficult and less realistic.

Physiological measures of workload are an alternative to subjective self-report methods. Several of these have the advantage of high temporal resolution [4]. In order to evaluate workload induced by in-vehicle systems, cognitive workload needs to be tested over rather short task durations. This requirement is met by EEG-based methods, pupillometry, and detection response tasks. Obviously, EEG-based methods are currently hardly suitable for widespread standard usability tests as the technical equipment is not only very expensive, but also time-consuming to set up and obstructive during testing. Pupillometry uses highly sensitive eye tracking equipment that can detect short-term changes in the pupil diameter. Changes in pupil dilation occur in direct response not only to changes in environmental illumination conditions but also to emotional and mental processes. Analyzing these characteristic bursts in pupil dilation can be a valid indicator of mental workload [5]. However, as alterations in lighting conditions or stimulus distance do also affect pupil size, filter mechanisms are needed to eliminate these influences. The Index of Cognitive Activity [6] corrects the pupil signal. Although this is a very innovative way of continuously measuring workload with extraordinarily high temporal resolution, the method and the underlying algorithms are still under validation in various research projects (e.g. [7]).

Detection response tasks (DRTs) are a promising method to evaluate cognitive workload [8-10]. The peripheral detection response task was developed by van Winsum, Martens and Herland [11] using visual stimuli. Subjects have to detect and respond to a stimulus that is presented regularly with slightly varying interstimulus intervals. The method is based on a dual task setting, in which the impairment in one task (the detection response task) is an indication of the workload imposed by another task (e.g. using a technical system). Although research shows that the DRT method is very sensitive to fluctuations of cognitive load, it is currently unclear as to how far different versions of the DRT, i.e. different stimulus presentation modalities, interact with different task modalities of the primary task. Attention and workload models like Wickens' multiple resource model [4] would predict that bottleneck effects are likely to occur if DRT presentation modality and primary task modality overlap. These effects would reduce the usefulness of the DRT method or at least would make adaptions necessary depending on primary task characteristics.

The present experiment compares effects of DRT presentation and secondary task modality in a controlled setting that allows for direct comparisons on the sensitivity of the DRT versions under resource conflicts. Three types of DRTs were evaluated: The standard visual remote peripheral detection response task (RDRT), the auditory detection response task (ADRT) and the tactile detection response task (TDRT). In order to evaluate the sensitivity of each of these DRTs, cognitively loading tasks were deployed in two levels of difficulty: easy (simpler variant with less task demand) and difficult (more complex variant with an increased task demand). This was intended to show whether the DRTs were sensitive enough to detect variations in workload. These cognitive tasks involved visual presentations, auditory presentations or no primary task presentation modality at all (purely cognitive task). The aim of this was to induce resource conflicts between for example the RDRT and the visually presented cognitive task in order to test for robustness of the DRTs.

2 Methods

2.1 Participants

Twenty four participants (12 female) took part in this experiment. The age range was between 21 and 42 years old, with a mean of 29 years (SD=5.17). All participants had normal or corrected-to-normal vision. Three of them were left handed.

2.2 Detection Response Tasks

The experiment was carried out in the Usability Lab at the BMW Group's Research and Innovation Center in Munich, Germany. Participants were seated at a desk with a laptop in front of them in central normal viewing distance. External loudspeakers were also located in front of the participants. All detection response tasks required participants to detect a stimulus and to respond to it via button press. The interstimulus interval randomly oscillated between 3000-5000 ms. Stimuli (auditory, visual or tactile) were presented for 1000 ms. Responses were always given by pressing a button (microswitch) that was attached to the index finger of the dominant hand (see Fig. 1, right side). Upon button press, the stimulus was switched off even if the presentation duration had not yet elapsed. Average reaction time and hit rate (responses within 2 s after signal onset) were recorded as performance indices.

Tactile Detection Response Task (TDRT). A vibrating cell phone motor was attached to the wrist of the non-dominant hand (see Fig. 1, left side). Prior to the beginning of the experiment, participants adjusted the vibration strength within a certain range to a comfortable level.

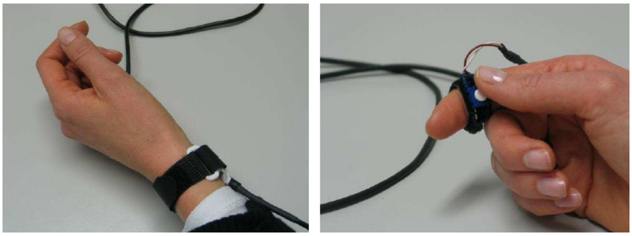

Fig. 1. Left side: Vibrating node attached to the wrist. Right side: Microswitch attached to the index finger of the dominant hand.

Remote Detection Response Task (RDRT). The RDRT was set at approximately 100 cm viewing distance. Four red LEDs were mounted on a black cardboard. These were arranged horizontally and spread symmetrically 11° and 23° from the center point as proposed by [12].

Auditory Detection Response Task (ADRT). Following the procedure of [10], the auditory stimulus was a 1 kHz sinus tone presented via loudspeakers. Participants were asked before the beginning of the experiment to adjust the audio volume within a certain range to a comfortable level.

2.3 Secondary Tasks

All secondary tasks were used to induce cognitive workload in two levels. In order to induce resource conflicts auditory task presentation and visual task presentation were introduced. Additionally, a purely cognitive task was part of the design in order to infer general sensitivity of the DRTs without resource conflict. The performance in the secondary tasks was monitored in order to see if participants were really engaged to a sufficient level.

Auditory N-Back Task. For auditory presentation mode, the n-back task [13] was used. It consists of the aural presentation of single digit numbers (0-9) with a system-paced interstimulus interval of 2.5 s. In the easy condition, the participants just had to repeat out loud each digit immediately after hearing it (0-back). In the difficult condition (2-back) participants needed to recall from memory the digit that was presented two digits before the currently presented numeric value and repeat it out loud while listening to further digit presented.

Visual N-Back Task. The procedure for visual presentation mode followed closely the auditory presentation mode. The n-back task was used with the same difficulty levels (0-back and 2-back). The digits were shown on the laptop monitor, each for 3 seconds without interstimulus interval.

Counting Task. A counting task was implemented that needed no further instruction during task execution than providing a three-digit initial number. Participants were requested to count upwards in steps of two in the easy condition and count downwards in steps of seven in the difficult condition. In this setting no visual or auditory processing load was present while executing the counting task.

2.4 Procedure

Prior to experimentation, participants went through a brief familiarization period in which all tasks were explained to them adequately by the experimenter. Participants adjusted signal levels of the ADRT and the TDRT, as described above. They were instructed to treat all tasks, primary task DRT and cognitively loading secondary task, with the same priority.

To minimize task switching effects, trials were arranged in DRT blocks. Participants performed the current DRT variant with all secondary tasks before starting the next DRT block. DRT variant order was randomized. Each secondary task was repeated three times. Participants were informed that the first trial always served as a training trial and was excluded from further analysis. The experiment took approximately one hour.

3 Results

The experiment was performed in order to evaluate the DRT as a workload assessment technique and to provide information on modality interference effects (resource conflicts) between primary and secondary task if these were based on the same presentation modality. Repeated measures ANOVAs were carried out on the data set for each secondary task in order to identify effects of difficulty level on reaction time and hit rate. Further t-test for each DRT within a secondary task scenario delivers information on the specific sensitivity of the DRT variant in detecting workload differences.

Results for the visual n-back task are shown in Fig. 2. Significant main effects for difficulty level were found for reaction time (F(1,23) = 32.68, $p < .05$) as well as for hit rate (F(1,23) = 13.48, $p < .05$). T-tests revealed that all of the DRTs proved to be effective and able to differentiate between workload levels. (reaction time: RDRT t(23) = -3.15, $p < .05$; TDRT t(23) = -4.72, $p < .05$; ADRT t(23) = -3.41, $p < .05$; hit rate: RDRT t(23) = 2.31, $p < .05$; TDRT t(23) = 2.90, $p < .05$; ADRT t(23) = 3.23, $p < .05$).

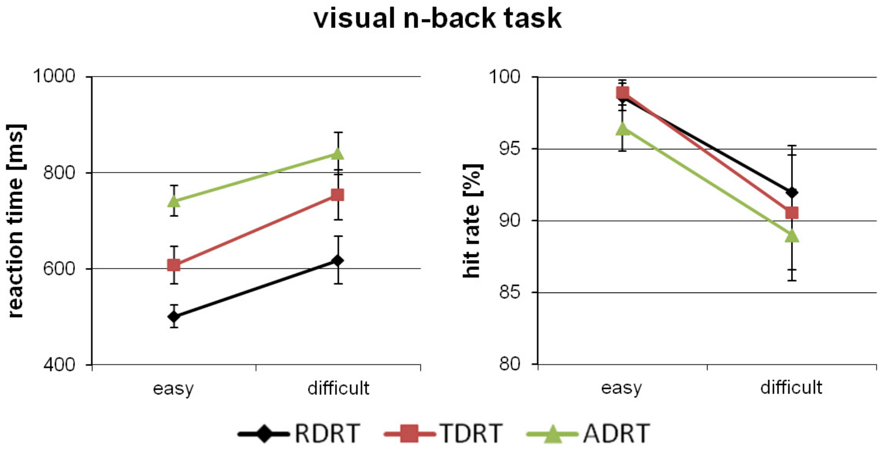

Fig. 2. Mean reaction times (left) and hit rates (right) for the DRTs during visual n-back task

For the auditory n-back task, a similar result pattern was found as can be seen in Fig. 3. Significant main effects for reaction time (F(1,23) = 129.45, p < .05) and for hit rate (F(1,23) = 19.55, p < .05) were confirmed by ANOVA. T-tests again showed significant differences for all DRTs while performing different levels of auditory n-back (reaction time: RDRT t(23) = -6.81, p < .05; TDRT t(23) = -6.67, p < .05; ADRT t(23) = -9.27, p < .05; hit rate: RDRT t(23) = 3.34, p < .05; TDRT t(23) = 4.13, p < .05; ADRT t(23) = 2.45, p < .05).

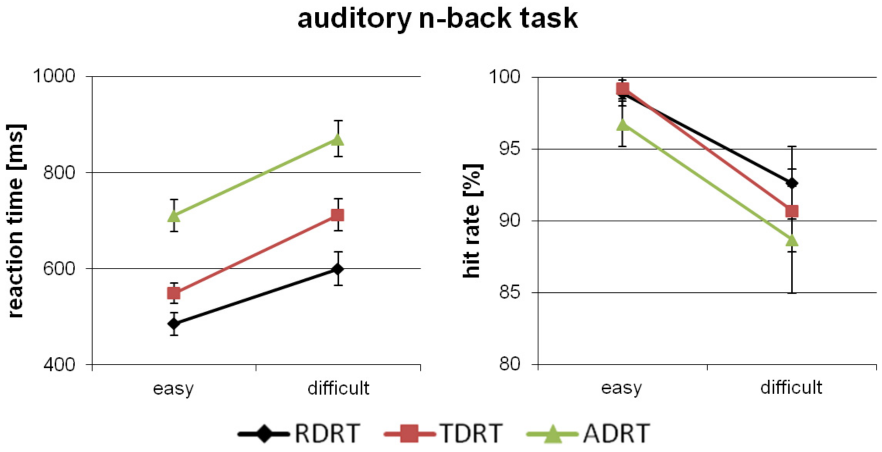

Fig. 3. Mean reaction times (left) and hit rates (right) for the DRTs during auditory n-back

The purely cognitive counting task led to slightly different results (see Fig. 4). While again a significant main effect for difficulty was found for reaction time (F(1,23) = 29.15, p > .05) and hit rate (F(1,23) = 18.44, p < .05), a significant interaction between difficulty and DRT variants on the hit rate dimension indicates differences in the degree of sensitivity of the DRTs in reaction to task levels (F(2,46) = 3.38, p < .05). As can be expected from Fig. 4, t-test analyses led to significant differences on reaction times (RDRT t(23) = -3.25, p < .05; TDRT t(23) = -3.44, p < .05; ADRT t(23) = -2.63, p < .05). When analyzing hit rate results, only TDRT (t(23) = 3.45, p < .05) and ADRT (t(23) = 3.18, p < .05) were able to discriminate between easy and difficult counting task. RDRT did not yield significant differences (t(23) = 1.92, ns).

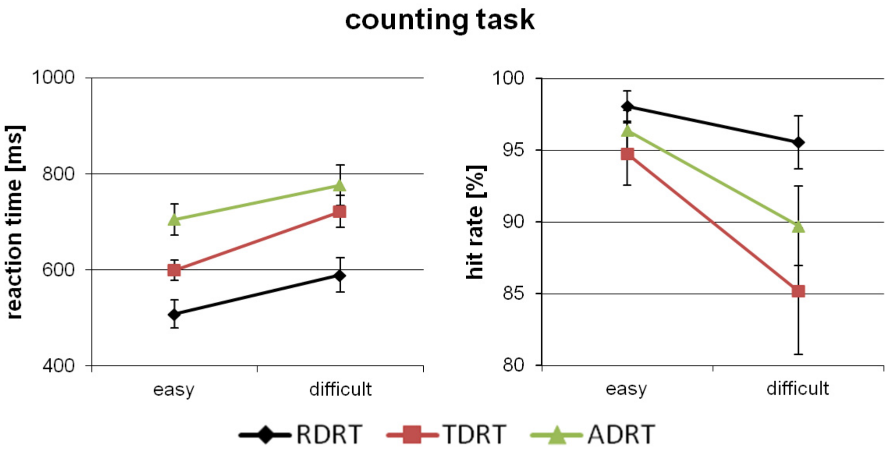

Fig. 4. Mean reaction times (left) and hit rates (right) for the DRTs during counting task

4 Discussion and Conclusion

The aim of the current study was to shed light on the sensitivity of different DRT variants while systematically varying secondary task presentation modality. Participants performed artificial cognitively loading tasks in different difficulty levels that were either presented aurally, visually or that consisted of a purely cognitive task without information presentation during task execution. DRTs under examination included visual remote peripheral, auditory and tactile stimuli detection.

The results strongly indicate that the DRT is a very robust and sensitive method to measure cognitive workload. Surprisingly, no interference effects were found between DRT stimuli presentation modality and secondary task modality. A possible explanation for this result can be that performing detection response tasks requires only little attention resources in itself. Both tasks could then be accomplished in spite of the resource conflict as no critical bottleneck level in task sharing was approached. An alternative explanation can be that the secondary task was cognitively loading but not requiring attentional resources in a level sufficiently high to induce critical modality interference. Future research thus needs to measure the amount of cognitive resource consumption during DRT execution only. Additional experiments should also systematically increase the level of sensory resource competition in the different secondary task modalities to check for interference effects.

All DRTs were basically able to discriminate between different levels of cognitive workload. The research presented here showed equally good results for ADRT, RDRT and TDRT when analyzing reaction times. As this is the dependent variable primarily used in DRT studies, the current study shows that researchers can choose the DRT modality most suitable for their experimental design in terms of avoiding resource conflicts without taking the risk of losing sensitivity. However, the experiment also provided evidence that the RDRT may not equally distinguish levels of workload when using hit rate as dependent variable. Future research should address the circumstances of this finding by focusing more strongly on providing comparable hit rate data in DRT studies.

Several steps need to be undertaken in order to establish DRTs as a suitable tool in driver distraction testing in the automotive context. First, more data is needed in how far DRTs can be used in triple task scenarios with concurrent driving simulation. Using a very simple driving task, [8] was able to show that tests participants are basically able to perform DRTs while driving and concurrently executing naturalistic and artificial secondary tasks. In a next step, data on more complex driving environments will be helpful. Second, a criterion needs to be defined which levels of DRT reaction time deterioration are critical when assessing cognitive driver distraction. Olsson and Burns [14] suggested that hit rates should not be less than 65% and reaction times should not fall below 800 ms. However, any threshold level needs to refer to concrete everyday driving situations in order to obtain ecological validity. A possible baseline for a socially accepted cognitive distraction level could be attention demanding conversations between driver and passenger.

Acknowledgments. The authors would like to acknowledge Otmar Hamm for his kind technical assistance as well as the Institute of Ergonomics team at the Technische Universität München (Klaus Bengler, Antonia Conti and Carsten Dlugosch) for their help in preparing this study and many fruitful discussions.

References

1. National Highway Traffic Safety Association (NHTSA), Department of Transportation: Visual-Manual NHTSA Driver Distraction Guidelines for In-Vehicle Electronic Devices. Report No. NHTSA-2010-0053, NHTSA (2012)
2. Zijlstra, F.R.: Efficiency in Work Behaviour. A Design Approach for Modern Tools. Delft University of Technology, Delft (1993)
3. Hart, S.G., Staveland, L.E.: Development of the NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. In: Hancock, P.A., Meshkati, N.M. (eds.) Human Mental Workload, pp. 139–183. North Holland B.V., Amsterdam (1988)
4. Wickens, C.D., Hollands, J.: Engineering Psychology and Human Performance. Prentice Hall, Upper Saddle River (2000)
5. Schwalm, M., Keinath, A., Zimmer, H.D.: Pupillometry as a Method for Measuring Mental Workload within a Simulated Driving Task. In: de Waard, D., Flemisch, F.O., Lorenz, B., Oberheid, H., Brookhuis, K.A. (eds.) Human Factors for Assistance and Automation, pp. 1–13. Shaker Publishing, Maastricht (2008)
6. Marshall, S.P., Davis, C., Knust, S.: The Index of Cognitive Activity: Estimating Cognitive Effort from Pupil Dilation. Technical Report ETI-0401. Eyetracking Inc., San Diego (2004)
7. Platten, F.: Analysis of Mental Workload and Operating Behavior in Secondary Tasks while Driving. Dissertation. Chemnitz University of Technology, Chemnitz (2012)

8. Conti, A., Dlugosch, C., Vilimek, R., Keinath, A., Bengler, K.: An Assessment of Cognitive Workload Using Detection Response Tasks. In: Stanton, N.A. (ed.) Advances in Human Aspects of Road and Rail Transport, pp. 735–743. Taylor & Francis Group, Boca Raton (2012)
9. Jahn, G., Oehme, A., Krems, J.F., Gelau, C.: 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 8, 255–275 (2005)
10. Merat, N., Jamson, A.H.: The Effect of Stimulus Modality on Signal Detection: Implications for Assessing the Safety of In-Vehicle Technology. Human Factors 50, 145–158 (2008)
11. van Winsum, W., Martens, M.H., Herland, L.: The Effects of Speech Versus Tactile Driver Support Messages on Workload, Driver Behaviour and User Acceptance. Report No. TM-99-C043, TNO Human Factors, Soesterberg (1999)
12. Martens, M.H., van Winsum, W.: Measuring Distraction: the Peripheral Detection Task. TNO Human Factors, Soesterberg (2000)
13. Mehler, B., Reimer, B., Coughlin, J.F., Dusek, J.A.: 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 (2009)
14. Olsson, S., Burns, P.C.: Measuring Driver Visual Distraction with a Peripheral Detection Task. Linköping University, Linköping (2000)