

DRIVER DISTRACTION THROUGH CONVERSATION MEASURED WITH PUPILLOMETRY

Carsten Dlugosch, Antonia S. Conti & Klaus Bengler
Technische Universität München, Institute of Ergonomics
Garching, Germany
Email: dlugosch@tum.de

Summary: Assessing a driver's mental workload during tasks that are not visual-manual is a challenging endeavor. Especially with the rapid development of speech systems, this is becoming increasingly important. Pupillometry promises to be a suitable physiological measurement method, sensitive to variations of cognitive workload. This driving simulator study shows that the pupillometry data indicate a significant increase in cognitive activity during conversation tasks regardless of the acoustic channel used.

INTRODUCTION

Distraction caused by secondary tasks has an impact on traffic safety (NHTSA, 2010). Up until recently, driver distraction was mainly considered as only pertaining to the visual-manual domain, which could be measured with various standardized methods such as eye tracking (ISO 15007, 2002) or occlusion (ISO 16673, 2007; Keinath et al, 2001). More recently, however, the driver is confronted with additional in-vehicle tasks that have little or no visual-manual component, such as voice control or conversation tasks. Hence, methods for measuring the mental workload or distraction of the driver became the focus of research. In literature, three different ways to measure mental workload can be found: Subjective, performance-based and physiological measurements. Subjective data reflect the perceived workload and are acquired by interviewing the subjects e.g. using NASA TLX (Hart & Staveland, 1988) or RSME (Zijlstra, 1993). Performance measurements determine the quality of the main or secondary task performance. With the driving task e.g. lane change task, LCT (ISO 26022, 2010), various metrics can be used to assess the distraction of a participant. In contrast, physiological methods evaluate the physical function of a participant, rather than his or her reactions e.g. heart rate (Mehler, Reimer, & Wang, 2011), EEG or pupillometry. The aim of this study was to evaluate different conversation types according to pupillometry.

RELATED RESEARCH

Conversation

Earlier studies have shown that voice tasks negatively influence the performance of other tasks (Atchley & Dressel, 2004). On road tests revealed that conversations deteriorated accurate response to traffic signs (McKnight & McKnight, 1993). Also the operation of voice-based e-mail systems has led to a worsening of reaction time and an increase in the subjectively measured mental workload (Lee, Caven, Haake, & Brown, 2001).

Detection Response Task

The *Detection Response Task (DRT)* is a representative of the performance measurement methods and has been used in many different studies (Bengler, Kohlmann, & Lange, 2012; Conti, Dlugosch, Vilimek, Keinath, & Bengler, 2012; Engström, Åberg, Johansson, & Hammarbäck, 2005; Jahn, Oehme, Krems & Gelau, 2005; Merat, Jamson & Kingdom, 2008). The participant responds to signals with the press of a button. The stimulus can be a visual, auditive or tactile signal. The stimulus is presented to the subject at random intervals every few seconds. Reaction time and hit rate allow concluding the mental workload.

Pupillometry

Pupillometry is a physiological method. It utilizes the changes of pupil diameter as a gauge of workload. Fluctuations of the pupil diameter are associated with fluctuations of workload; the more difficult a task, the larger the pupil diameter (Kahneman & Beatty, 1966; Beatty & Wagoner, 1978). A major advantage to employing pupil metrics is that such a measure does not disrupt or require the addition of tasks for its measurement. However, this method is also susceptible to negative characteristics, such as, sensitivity to environmental luminance and emotional arousal (Iqbal, Zheng, & Bailey, 2004). There are different metrics used to infer the mental load like mean pupil diameter or mean pupil diameter change rate (Palinko, Kun, Shyrovkov, & Heeman, 2010). These metrics do not rely on highly precise pupil measurements and are therefore suitable for remote eye trackers. As the pupil size is highly dependent on environmental factors (e.g. pupillary light reflex), Sandra Marshal (2002) introduced the *index of cognitive activity (ICA)*, a metric based on fast, temporary pupil dilation that seem to have a strong correlation to mental stress and is largely independent of the effects of light. Due to improvements in camera technology and data analysis, it is becoming increasingly easy to use and interesting for usability studies (Schwalm, 2009).

METHOD

Overview

In a static simulator the subjects drove along a simple track and simultaneously performed a number of additional (secondary) tasks while wearing a head mounted eye tracker for pupil measurement. In addition, the subjects also performed a detection response task, to get performance based data. Each experimental session lasted around 30 minutes.

Subjects

Twenty-four persons participated in this study. Since pupillometry was not possible for every subject (e.g. too thick eyelashes or too small lid opening), only 18 subjects were evaluated (8 females and 10 males), aged between 18 and 30 years old ($M = 24.4$, $SD = 4.2$). All participants held a valid driver's license. Participants wearing glasses were excluded beforehand to prevent pupillometry artifacts caused by reflections of the screen.



Figure 1. Experimental setup: Participant is performing driving task, head-mounted detection response task and SuRT at the same time

Driving task

The experiment was carried out in a static driving simulator (SILAB, WIVW GmbH), equipped with an authentic driver's seat and original steering wheel. The simulated driving scene was presented on a plasma screen. No rear view or mirrors were used. The driving scene was a highway with two lanes in each direction with straight and curved sections. The task was to follow a red leading car at the right lane at a distance of 50 m. For the ease of estimation, roadside poles were positioned every 50 m. The leading vehicle drove at a constant speed of about 72 km/h. There were no additional vehicles on the road.

Pupillometry

For the pupillometry measurement, a head mounted eye tracker (EyeLink II, SR Research) was used. The original mount was replaced by a baseball-type cap to improve wearing comfort. The ICA was calculated by EyeWorks (EyeTracking, Inc). The values were z-normalized for each participant.

Detection Response Task

During the experiment the participants performed a head mounted detection response task. A single red LED was mounted to the baseball cap. The LED was viewed at a distance of 18 cm. The signal was presented randomly every 3000 – 4000 ms for 1000 ms. Participants responded by pressing a button on their left index finger. The results for the DRT can be found in Conti, Dlugosch, Schwarz & Bengler (2013).

Secondary Tasks

In addition to the driving task and the detection response task the subject performed several secondary tasks: Two artificial tasks, three natural conversational tasks and three mainly visual manual tasks were performed. The two artificial tasks, auditory n-back and SuRT, were included as references and for better comparison with other laboratories participating in the DRT standardization group ISO/NP WD 17488.

During the *auditory n-back* task, numbers were read slowly by a computer to the participant. After each number, the subject has to recall and repeat a number that was previously stated. In this experiment, the number said two steps prior to the current number was to be repeated ($n = 2$). This task is cognitively demanding, rather than visually. (Mehler, Reimer, and Dusek, 2011)

The *SuRT (Surrogate Reference Task)* is a visual search task requiring visual-manual interaction. On a black screen with many white circles, the subject is to find the one smaller circle and select it by using cursor keys on a small keyboard. This task was used in its difficult variant i.e. the difference between the size of the targets and the size of the distractors was marginal. (ISO 14198, 2011)

Three conversation scenarios were investigated: one with a passenger, and two with a simulated hands-free cell phone conversation, once with a clear, once with a noisy connection. All conversations were performed by the same confederate (female, 26 years old). The content was always small talk and the subject was able to change the topic at any time. It has been tried to achieve the most balanced communication.

For the *passenger conversation task*, the interlocutor was positioned next to the participant similar to an actual in vehicle situation. For both *clear* and *noisy phone conversation* tasks, a hands-free cell phone was simulated. In the noisy condition, the quality of transmission was compromised and simulated a poor cell phone connection. Specifically, the voice of the driver's conversation partner was distorted, band pass filtered and superimposed with a background noise. Moreover, the connection was frequently interrupted to provoke communication difficulties and to force follow up questions and increase dialogue.

Furthermore, three mainly visual-manual tasks were addressed: sorting of candy, searching in a handbag and operating a touchscreen. These are beyond the scope of this paper and will not be discussed.

Procedure

The tasks were first presented in a multimedia presentation. Afterwards each task was rehearsed (except the conversation) and outstanding issues were clarified. It also was pointed out that it is the main priority to drive safely. The tasks were carried out in two blocks: In one block the actual study was carried out (driving task, DRT and secondary task at the same time). The other block consists of the baseline tasks. Half of the subjects began with the baseline block. Within the blocks tasks were permuted. The artificial scenes lasted one minute each, the conversation

tasks 2-3 minutes depending on willingness to communicate. For the artificial tasks the entire period (1 minute) was analyzed and for the conversations only the last minute was considered.

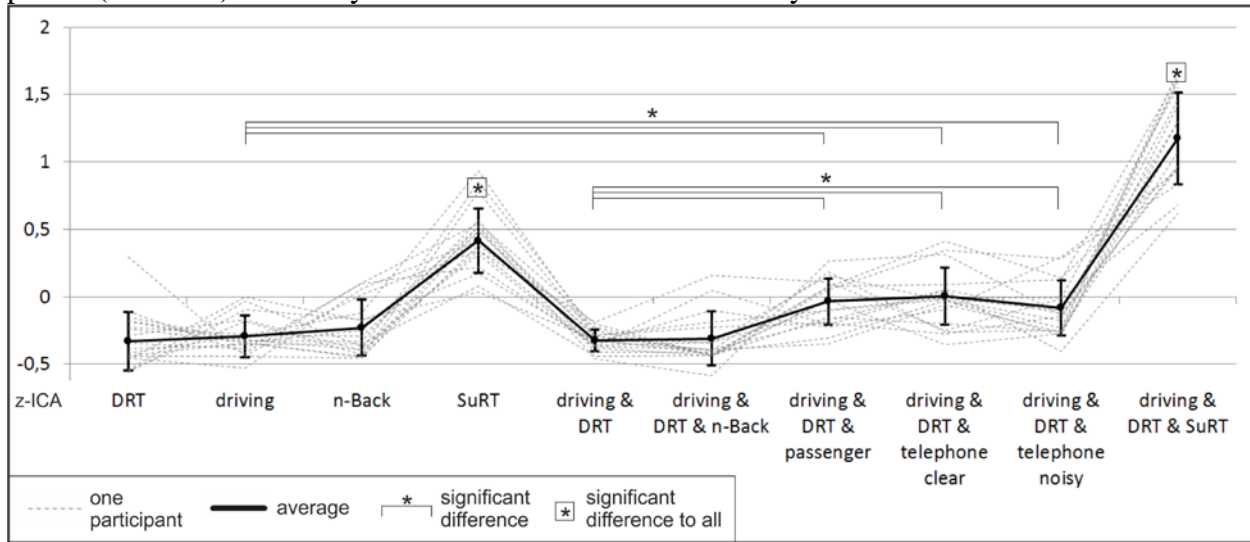


Figure 2. Mean z-ICA (+/- 1 SD, N = 18) shows an increase in cognitive load during communication, and greater during the visual-manual task

RESULTS

The mean values and standard deviation of the normalized ICA for the different conditions are listed in table 1. One repeated measures ANOVA was performed to assess the differences in mean ICA values across task conditions. Mauchly’s test indicated that the assumption of sphericity had been violated, $\chi^2(44) = 70.70, p < .01$, therefore the Greenhouse-Geisser correction ($\epsilon = .50$) was used. The mean ICA values significantly differed depending on task conditions, $F(4.49, 67.39) = 72.07, p < .001, \eta^2 = .83$. The Bonferroni post hoc test revealed that a significant difference was found between the SuRT and all other conditions (driving + DRT + telephone noisy, $p < .01$; driving + DRT + telephone clear, $p < .05$; others, $p < .001$) as well as between driving + DRT + SuRT and all other conditions (driving, $p < .01$; others, $p < .001$). There is also a difference between each of the both driving baseline conditions and each of the three conversation tasks ($p < .01$). Figure 2 visualizes these findings.

DISCUSSION

The experiment evaluated different conversation types according to pupillometry. All three conversation tasks showed a significant increase of ICA value, reflecting the additional mental workload induced by the tasks. The different conversation tasks were not discriminated, which is in line with other studies that showed no difference depending on the physical presence of the interlocutor (Nunes et al., 2002).

Table 1. Mean and standard deviation for the normalized ICA values (N = 18)

condition	Baseline				driving & DRT	driving + DRT +				
	DRT	driving	n-Back	SuRT		n-Back	passenger	phone clear	phone noisy	SuRT
mean zICA	-.330	-.293	-.229	.415	-.325	-.309	-.035	.005	-.081	1.173
std. dev.	.217	.154	.205	.236	.079	.199	.169	.209	.204	.341

The SuRT yields a very high ICA values, even more in combination with the driving task and DRT. Similar results can be found in Schwalm (2009). Furthermore, similar results were found in the DRT performance of the experiment (Conti et al, 2013).

ACKNOWLEDGMENT

Parts of this study were funded by BMW Group Research and Technology, Munich, Germany.

REFERENCES

- Atchley, P., & Dressel, J. (2004). Conversation Limits the Functional Field of View. *Human Factors*, 46(4), 664–673.
- Beatty, J. & Wagoner, B.L. (1978). Pupillometric signs of brain activation vary with level of cognitive processing. *Science*, 199, 1216-1218.
- Bengler, K., Kohlmann, M., & Lange, C. (2012). Assessment of cognitive workload of in-vehicle systems using a visual peripheral and tactile detection task setting. In: *Work: A Journal of Prevention, Assessment and Rehabilitation 41* (Supplement 1/2012), S. 4919–4923.
- Conti, A., Dlugosch, C., Schwarz, F., & Bengler, K. (2013). Driving and speaking: Revelations by the head-mounted detection response task. *Proceedings of the Seventh International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design*, Bolton Landing, New York.
- Conti, A., Dlugosch, C., Vilimek, R., Keinath, A., & Bengler, K. (2012). An assessment of cognitive workload using detection response tasks. *Proceedings of the Fourth International Conference on Applied Human Factors and Ergonomics*, San Francisco, California.
- Engström, J., Åberg, N., Johansson, E., & Hammarbäck, J. (2005). Comparison between visual and tactile signal detection tasks applied to the safety assessment of in-vehicle information systems. *Proceedings of the Third International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design*, Rockport, Maine, 232-239.
- Hart, S. G., & Staveland, L. E. (1988) Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In P. A. Hancock & N. Meshkati (Eds.), *Human Mental Workload*. Amsterdam: North Holland Press.
- Iqbal, S., Zheng, X., & Bailey, B. Task-evoked pupillary response to mental workload in human-computer interaction. *CHI'04 extended abstracts on Human factors in computing systems*, ACM (2004), 1477-1480.
- ISO 14198 (2011). Road vehicles - Ergonomic aspects of transport information and control systems - Calibration tasks for methods which assess demand due to the use of in-vehicle systems. Revised Draft Version.
- ISO 15007 (2012). Road vehicles - Measurement of driver visual behaviour with respect to transport information and control systems.
- ISO 16673 (2007). Road vehicles - Ergonomic aspects of transport information and control systems - Occlusion method to assess visual distraction due to the use of in-vehicle systems.

- ISO 26022 (2010). Road vehicles - Ergonomic aspects of transport information and control systems - Simulated lane change test to assess in-vehicle secondary task demand.
- Jahn, G., Oehme, A., Krems, J., & Gelau, C. (2005). Peripheral detection as a workload measure in driving: Effects of traffic complexity and route guidance system use in a driving study. *Transportation Research Part F: Traffic Psychology and Behaviour*, 8(3), 255-275.
- Kahneman, D. & Beatty, J. (1966). Pupil diameter and load on memory, *Science*, Vol 154(3756), 1966, 1583-1585. doi: 10.1126/science.154.3756.1583
- Keinath, A., Baumann, M., Gelau, C., Bengler, K., & Krems, J. F. (2001). Evaluation of in-vehicle HMI using occlusion techniques: Experimental results and practical implications. *Exploring the Occlusion Technique: Progress in recent research and applications*, Turin, Italy.
- Lee, J. D., Caven, B., Haake, S., & Brown, T. L. (2001). Speech-based Interaction with In-vehicle Computers: The Effect of Speech-based E-mail on Drivers' Attention to the Roadway. *Human Factors*, 43(4), 631-40.
- McKnight, A.J., & McKnight, A.S. (1993). The effect of cellular phone use upon driver attention. *Accident; analysis and prevention*, 25(3), 259-65.
- Marshall, S. (2002). The Index of Cognitive Activity: measuring cognitive workload. In: J. J. Persensky, B. Hallbert und H. S. Blackman (Hg.): *Proceedings of the 2002 IEEE 7th Conference on Human Factors and Power Plants*. IEEE 7th Conference on Human Factors and Power Plants. Scottsdale, 15.-19. September 2002. IEEE. New York, S. 7-5-7-9.
- Mehler, B., Reimer, B. & Dusek, J.A. (2011). . MIT AgeLab White Paper Number 2011-3B. MIT AgeLab Delayed Digit Recall Task (n-Back), Massachusetts Institute of Technology, Cambridge, MA.
- Mehler, B., Reimer, B. & Wang, Y. (2011). A Comparison of Heart Rate and Heart Rate Variability Indices in Distinguishing Single Task Driving and Driving Under Secondary Cognitive Workload. *Proceedings of the 6th International Driving Symposium on Human Factors in Driver Assessment, Training, and Vehicle Design*, Lake Tahoe, CA. 590-597.
- Merat, N., Jamson, A. H., & Kingdom, U. (2008). The Effect of Stimulus Modality on Signal Detection: Implications for Assessing the Safety of In-Vehicle Technology. *Human Factors*, 50(1), 145-158.
- Nunes, L., & Recarte, M. A. (2002). Cognitive demands of hands-free-phone conversation while driving. *Transportation Research Part F: Traffic Psychology and Behaviour*, 5(2), 133-144.
- NHTSA - National Highway Traffic Safety Administration (2010). *Distracted Drive Report released September 2010*. DOT HS 811 379
- Palinko, O., Kun, A. L., Shyrovkov, A., & Heeman, P. (2010), Estimating Cognitive Load Using Remote Eye Tracking in a Driving Simulator, *ETRA 2010*
- Schwalm, M. (2009). Pupillometrie als Methode zur Erfassung mentaler Beanspruchungen im automotiven Kontext. Dissertation. Universität Saarbrücken, 2009.
- Zijlstra, F.R.H. (1993). Efficiency in work behavior. A design approach for modern tools. PhD thesis, Delft University of Technology. Delft, The Netherlands: Delft University Press.