

DRIVING AND SPEAKING: REVELATIONS BY THE HEAD-MOUNTED DETECTION RESPONSE TASK

Antonia S. Conti¹, Carsten Dlugosch¹, Felix Schwarz² & Klaus Bengler¹

¹Technische Universität München, Institute of Ergonomics, Garching, Germany

²BMW Group Research and Technology, Munich, Germany

Email: conti@lfe.mw.tum.de

Summary: The cognitive workload of speech-related activity needs to be examined in an economic and simple way. This is especially important as in-vehicle technology is becoming more cognitive with, for example, the use of speech-interaction and industry will need a way to keep pace with new technologies. One proposed way to measure cognitive workload is the detection response task (DRT) method. In this study, the DRT was used to assess different speech-related cognitive tasks. Three conversation tasks and the n-back task were performed together with a simulated driving task and a head-mounted DRT (HDRT). The aim was to evaluate the conversation and n-back tasks with the HDRT and to quantify the respective cognitive workload. Results show an increase in HDRT reaction times when additional cognitive tasks are performed relative to baseline measurements. In line with other research methods, the HDRT provided a reliable measurement of additional workload.

INTRODUCTION

Previous research has found that speech-related activity, including speech-interaction and conversation, can negatively affect additional task performance; for example, decrements in peripheral detection and discrimination tasks (Atchley & Dressel, 2004) or change-detection (McCarley et al., 2004). Similar findings have been reported in-vehicle. McKnight and McKnight (1993) found that when participants were engaged in a conversation task, their ability to appropriately respond to traffic scenarios worsened. Additionally, Lee, Caven, Haake, and Brown (2001) found that both driver RTs and subjective workload increased when drivers used a speech-based e-mail system.

Effects of speech-interaction can also be found remotely when an interlocutor is not physically present; often these tasks are made to resemble hands-free cell phone usage. Remote conversation tasks are associated with a disruption of visual attention and an increase of traffic signal misses and RTs (Strayer & Johnston, 2001; Strayer, Cooper, & Drews, 2004). However, it has also been shown that the location of the interlocutor has no real distracting influence; rather, conversation itself is the main source of distraction (Nunes & Recarte, 2002; Recarte & Nunes, 2003). Amado & Ulpinar (2005) also found no difference between different conversation types where an interlocutor was either physically present or not.

Since interaction by speech or conversing is a cognitive activity, different than visual-manual actions, cognitive workload needs to be measurable. The detection response task (DRT) method is gaining momentum as a method to quantify cognitive or attentive workload (Bengler, Kohlmann, & Lange, 2012; Conti, Dlugosch, Vilimek, Keinath & Bengler, 2012; Engström, Åberg, Johansson, & Hammarbäck, 2005; Jahn, Oehme, Krems & Gelau, 2005; Merat, Jamson

& Kingdom, 2008) and is currently in the process of standardization (ISO/NP WD 17488). The DRT is used according to the secondary task method where one task (e.g. the DRT) is performed in addition to other tasks as a gauge of their workload (Ogden, Levine & Eisner, 1979; Waard, 1996; Wickens & Hollands, 2000). Participants are continuously presented with a stimulus (either visual, auditory, or tactile) that turns on at random intervals. As soon as a participant perceives the stimulus (ex. the LED turning on or motor vibrating) they indicate this by pressing a button. The performance of the DRT task, reaction times (RTs) and misses, can be interpreted as the degree to which the other tasks required cognitive or attentional resources.

The aim of the current experiment was to evaluate different conversation types according to the DRT. Furthermore, this experiment also tested whether the DRT could replicate previous findings that conversation tasks impair additional task performance. Three different natural conversation types were used in this experiment. The driver's interlocutor was positioned either next to the driver, as a conversation with a passenger, or remotely in a separate room, as a cell phone conversation. Cell phone connection quality was also manipulated as either clear or not. These manipulations of the conversation task were to simulate realistic in-vehicle occurrences and to evaluate whether clarity would influence concurrent DRT performance. The n-back task (Mehler, Reimer, Coughlin & Dusek, 2009) was also included as the common barometer used by all ISO collaborating laboratories. The n-back task also served as the artificial cognitive task to which the naturalistic conversation tasks could be compared with.

METHODS

Participants

Twenty-four persons participated in this study. Two participants were excluded: one was red-green color blind and the other due to a computer malfunction in the data recording phase. The resulting 22 participants, 10 females and 12 males between the ages of 18 and 33 ($M = 25$, $SD = 4.43$), qualified for analysis. All participants held a valid driver's license. Participants reported to have normal or corrected-to-normal vision and all, except for one participant, were right handed.

Tasks and Equipment

The current experiment included two main tasks: head-mounted DRT (HDRT), simulated driving, and two cognitive tasks: n-back and conversation (3 levels). Additional visual and visual-manual tasks in addition to pupil metrics were also tested in this experiment, though not presented here (for pupil metrics, see Dlugosch, Conti, & Bengler, 2013). The HDRT was used in this experiment due to its more consistent performance (Conti et al., 2012). Participants wore a head-mounted LED to facilitate the detection response task. A single red LED was mounted to a baseball-type cap. The LED was viewed at a distance of 18 cm, measured from the cap's brim joint. The interstimulus interval randomly varied between 3000 – 4000 ms and stimuli were presented for 1000 ms. Participants responded to the HDRT by pressing a button fixed to their left index finger against the steering wheel and were instructed to do so as quickly and accurately as possible.

The simulated driving task was a four-lane highway scene (2 lanes per direction) with a red leading car travelling at approximately 72 km/h (simulator software: SILAB; Veitshöchheim, Germany). The vehicle was automatic and required no gear shifting. Participants were instructed to drive in the center of the right-hand lane and to follow the car at a distance of 50 meters. This distance was approximately the distance between two highways markers, which were also included in the highway scenery on the right side of the road.

The cognitive tasks were employed to manipulate the participants' cognitive workload. The n-back task was used as $n = 2$. In this task, numbers were dictated to a person and the participant repeated the number said two steps prior to the current number. The conversation tasks were divided in to three: with a passenger, over a hands-free cell phone with a clear connection, and with a noisy connection. The same confederate was the interlocutor for all conversations within and across participants. For the passenger conversation, the confederate sat next to the participant. For the cell phone conversations, the confederate communicated via headset from a room adjacent to the laboratory. The "noisy" cell phone task involved distorting the confederate's voice and adding a specified "shopping mall" background noise (software: MorphVox Pro; Screaming Bee Inc.; Palm Coast, FL). A 250-2000 Hz band pass filter was also applied. The "noisy" manipulation was the same for all participants. Additionally, often when the confederate would say a word critical to the meaning of the sentence, the audio signal was briefly muted (ex. "I met John at the cinema" versus "I met ---n at the ---nema). This was done to simulate a very poor cell phone connection. Participants were instructed to converse as they normally would and that they were free to change the topic at any time. No specific discussion topics were required and the conversations typically resembled small talk. It was ensured that participants were continually engaged in this task for the duration of the measurement.

Procedure

Experiments were carried out in a fixed simulator with a centrally positioned driver's seat. Participants first watched a multimedia instructional presentation, where they were introduced to each of the tasks. After each task was explained, participants were able to practice the task and ask questions. Since conversing is a natural activity, the conversation tasks were not practiced. The instructional presentation ended with the overall instruction to perform the simulated driving task as the primary task; no additional specific prioritization regarding the other tasks was given.

Afterwards, participants began the experiment with either a baseline block (including a static baseline, without the simulated driving task, or dynamic baseline, with it – each performed for 1 minute) or an experimental block. Participants performed both blocks once during the experimental session. Each block consisted of a certain number of scenes or conditions that specified the tasks to be performed. In the experimental block, the HDRT and simulated driving task were always performed together with each of the cognitive tasks (4 scenes total, reflecting the number of cognitive tasks). Scenes with the n-back task were performed for 1 minute, while those with the conversation tasks were allowed to run for up until 2 minutes or more. Participants were always told by the experimenter which cognitive task they were to perform in the upcoming scene. Both block and scene order was randomized across participants.

RESULTS

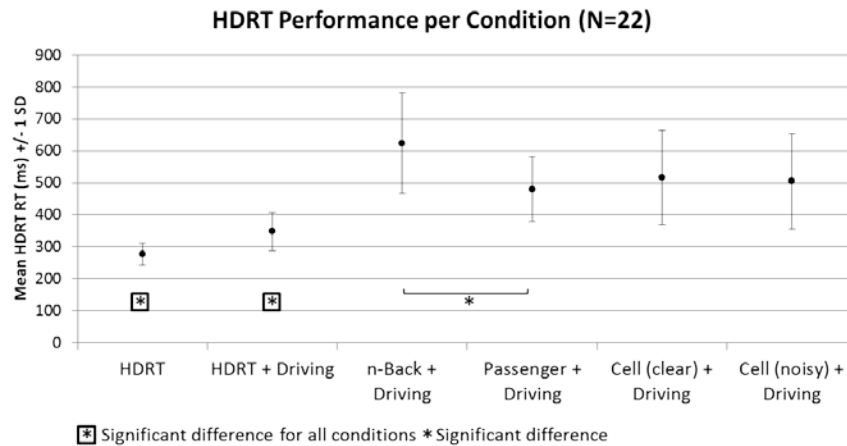


Figure 1. Mean HDRT RT (ms) +/- 1 SD per condition (N = 22). HDRT & HDRT + Driving served as baselines; static and dynamic, respectively. All other conditions were performed as a triple task condition with the listed tasks plus the HDRT

Only HDRT RT data ranging from 100 – 2000 ms post stimulus onset were considered. All scenes except for the conversation conditions were analyzed for their entire duration of 1 minute. For the conversation tasks the last minute was assessed in order to allow participants time to engage in the conversation. In Figure 1, the RTs (ms) across conditions can be found for the HDRT. For all conditions the hit rates were above 90%. In terms of RTs, it can be seen that the HDRT performed as static yielded lower RTs than when performed as dynamic. With the addition of a cognitive task, RTs increase, especially when the n-back task is performed. RT means and standard deviations can be found in Table 1.

One repeated measures ANOVA was performed to assess the HDRT's discrimination of different conditions in terms of mean DRT RT. Mauchly's test indicated that the assumption of sphericity had been violated, $\chi^2(14) = 45.80, p < .001$, therefore the Greenhouse-Geisser correction ($\epsilon = .65$) was used. The mean HDRT RTs were significantly different depending on task conditions, $F(3.23, 67.70) = 40.29, p < .001, \eta_p^2 = .66$. The Bonferroni post hoc test revealed a significant difference between the static baseline and all dynamic conditions HDRT RTs, $p < .001$. The dynamic baseline was also significantly different from all conditions where additional cognitive tasks were performed, $p < .001$. The mean HDRT RTs while performing the simulated driving and the n-back tasks were significantly different from the passenger conversation task condition, $p < .001$, however, not significant for clear and noisy cell phone conversation conditions. No significant differences were found for the HDRT RTs across the conversation task manipulation.

DISCUSSION

The aim of the current experiment was to evaluate different conversation types according to the DRT. The HDRT was evaluated to see if additional cognitive workload could be detected in its performance. The n-back task was included as an artificial, cognitive task to which the HDRT performance during natural tasks could be compared and to be able to compare experimental settings across collaborating laboratories.

Table 1. Mean HDRT RT (ms) for all conditions (N=22)

Conditions	Baseline	Baseline	HDRT + Driving +			
	HDRT	HDRT + Driving	n-Back	Passenger	Cell (clear)	Cell (noisy)
Mean RTs	276.69	350.63	624.36	478.47	519.40	516.79
Std. Deviation	34.26	60.05	158.27	100.99	147.57	148.59

The results of this study indicate that the HDRT is sensitive to additional cognitive workload and able to discriminate between certain conditions. According to the HDRT, conditions with additional cognitive tasks yield longer RTs and can be interpreted as more cognitively demanding than conditions without such additional tasks. Specifically, it was found that DRT RTs were longer when participants conversed or performed the n-back task relative to DRT baselines. The HDRT, however, did not discriminate between the different types of conversation, both in terms of interlocutor location and clarity of the cell phone connection. According to the HDRT, it seems that there is no substantial difference between the conversation types used in this study. In line with previous reports (Amado et al., 2005; Nunes et al., 2002; Recarte et al., 2003), the finding that the position of the interlocutor does not affect additional task performance (viz. DRT performance) is supported, suggesting that conversing itself is the primary source of workload.

A useful contribution of this experiment is that it allows a comparison of DRT values for both a naturalistic, acceptable-to-perform in-vehicle task (viz. conversing), and that of an artificial cognitive task (viz. n-back). It was seen that DRT performance worsened when concurrent n-back performance was required, relative to the baselines and to the passenger conversation task. Future research should expand on this to determine whether this finding is due to a possible limitation of the DRT method or rather to the conversation task itself. Since the aim of this paper was to evaluate speech-related activity with the DRT, driving performance data was not considered. In order to derive effects on traffic safety, subsequent analyses of this data will consider the simulated driving task performance and its relation to the DRT data. Possible cross-modal interferences with the DRT under such experimental settings should also be considered.

CONCLUSIONS

The present experiment evaluated passenger and remote conversation types while performing a simulated driving task, with the HDRT. The n-back was included as a barometer of comparison across coordinating laboratories and to compare the natural conversation tasks to an artificial task. The HDRT was sensitive to the different cognitive tasks; however, the different conversation types were not discriminated by to the HDRT. Here, the HDRT was able to replicate previous findings that concurrent conversation tasks are associated with a decrease in performance of other tasks and that the location of an interlocutor does not additionally affect DRT performance. It can be concluded that the HDRT can be used to assess general, speech-related activity and facilitates the comparison of different tasks in the same measurement terms.

ACKNOWLEDGMENTS

This project was funded by BMW Group Research and Technology, Munich, Germany.

REFERENCES

- Amado, S., & Ulupinar, P. (2005). The effects of conversation on attention and peripheral detection: Is talking with a passenger and talking on the cell phone different? *Transportation Research Part F: Traffic Psychology and Behaviour*, 8(6), 383–395.
- Atchley, P., & Dressel, J. (2004). Conversation Limits the Functional Field of View. *Human Factors*, 46(4), 664–673.
- 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., 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.
- Dlugosch, C., Conti, A., & Bengler, K. (2013). Driver distraction through conversation measured with pupillometry. *Proceedings of the Seventh International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design*, Bolton Landing, New York.
- 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.
- ISO/NP WD 17488: Road vehicles - Transport information and control systems - Man machine interface, 2012-02, International Organization for Standardization, Geneva, Switzerland. http://www.iso.org/iso/catalogue_detail.htm?csnumber=59887
- 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.
- 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.
- McCarley, J. S., Vais, M. J., Pringle, H., Kramer, A. F., Irwin, D. E., & Strayer, D. L. (2004). Conversation disrupts change detection in complex traffic scenes. *Human factors*, 46(3), 424–36.
- McKnight, A.J., & McKnight, A.S. (1993). The effect of cellular phone use upon driver attention. *Accident; analysis and prevention*, 25(3), 259–65.
- 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., & 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.
- Ogden, G. D., Levine, J. M., & Eisner, E. J. (1979). Measurement of workload by secondary tasks. *Human Factors*, 21, 529–548.
- 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(2), 119–137.
- Strayer, D. L., & Johnston, W. A. (2001). Driven to distraction: Dual-task studies of simulated driving and conversing on a cellular phone. *Psychological Science*, 12, 462-466.
- Strayer, D. L., Cooper, J. M., & Drews, F. A. (2004). What do drivers fail to see when conversing on a cell phone? *Proceedings of the Human Factors and Ergonomics Society 48th Annual Meeting* (pp. 2213–2217).
- Waard, D. de (1996). The measurement of drivers' mental workload. PhD Thesis, University of Groningen.
- Wickens, C. D., & Hollands, J. (2000). *Engineering Psychology and Human Performance*, 3rd edn. (Upper Saddle River, NJ: Prentice Hall).