The Effect of Task Set Instruction on Detection Response Task Performance

Antonia S. Conti, Carsten Dlugosch, & Klaus Bengler
Institute of Ergonomics, Technische Universität München, Garching, Germany

Abstract

Detection response tasks (DRTs) have been suggested as a way to measure online levels of cognitive workload. DRTs have been used in various applied settings under different task instructions. Especially in situations where multiple tasks are performed simultaneously (e.g. a driver operating a vehicle and using a navigation device), how someone allocates attention and other cognitive resources is also based on the given task instruction. Task instruction is important and can affect the strategy one takes on to complete a task, affecting the metric used to gauge task performance (e.g. reaction times). This, of course, can affect the results and the conclusions drawn based on these results; which, in an applicative setting, has real-life outcomes (e.g. law making, car manufacturers, etc.). As the DRT method is in the process of being standardized, the effect of various instruction types on task performance is to be investigated. The present paper reports on a between-group study where participants performed a triple-task scenario, involving a simulated driving task, under four different instructions. No difference in mean DRT RTs was found according to task instruction, suggesting that the complexity of a triple-task condition subjugates any influence that different task instructions could have on the performance metric.

Introduction

Varieties of detection response tasks (DRT) have been implemented for many years in driver distraction research (e.g., Bengler et al., 2012; Conti et al., 2012; Engström et al., 2005; Jahn et al., 2005; Merat & Jamson, 2008; Olsson & Burns, 2000; van Winsum et al., 1999). Over time, this method has matured and is currently being used to measure the attentive effects of cognitive load (Bruyas & Dumont, 2013; Conti et al., 2013; Engström et al., 2013; Harbluk et al., 2013; Young et al., 2013). The DRT presents continuous stimuli (visual, auditory, or tactile) to which a participant responds via button press. The performance metrics of the DRT, reaction times (RTs) and misses, can be interpreted as the degree to which the other tasks required cognitive or attentional resources. The DRT is used according to the secondary task method (Knowles, 1963; Ogden et al., 1979; De Waard, 1996; Wickens & Hollands, 2000) and is performed in addition to other tasks.

In D. de Waard, K. Brookhuis, R. Wiczorek, F. di Nocera, R. Brouwer, P. Barham, C. Weikert, A. Kluge, W. Gerbino, & A. Toffetti (Eds.) (2014). Proceedings of the Human Factors and Ergonomics Society Europe Chapter 2013 Annual Conference. ISSN 2333-4959 (online). Available from http://hfes-europe.org

As part of the progress of the DRT method, the DRT is being standardized under ISO/CD 17488. Essential to this process is the standardization of the instructions to be used. Especially because the DRT is performed in addition to other tasks, a standardized task set instruction is needed to be able to advise the participant on how to prioritize task performance when they are required to perform multiple tasks at a time.

Instructions for an experiment are typically standardized with the intent to optimally orient participants to an experiment. In addition to serving as an introduction and explanation, it is reasonable to consider that the specific instructions a participant receives can influence his or her task performance. However, previous research on this topic both support and oppose this.

For example, Haider and Frensch (1999) found that task performance metrics, reaction times (RTs) and errors, as well as the extent of information processing, were sensitive to different instructions (participants were told to optimize accuracy, speed, or a combination of the two). The authors concluded that this finding could be due to the setting of a response criterion based on the given instructions. Task instructions have also been observed to influence the mental representations of task responses, such that the instructed response labels (colours and location, in this case) in a dual-task setting affect the coding of responses and task performance (Wenke & Frensch, 2005). Trottier and Pratt (2005) found that saccadic RTs were reduced when the task instruction was to gather information on a scene (an instruction relevant to the function of a saccade), rather than to be as fast as possible. Another performance modulation due to instructions was reported by Ansorge and Newmann (2005), where priming effects were eliminated by instructing the prime as task irrelevant.

On the contrary, the effect of task instruction has also been reported to have a minimal effect task performance. In 2001, Levy and Pashler evaluated whether dual-task slowing could be altered through instructions that varied task priority. They found that manipulating task instructions did not alone suffice to eliminate dual-task interference. Additionally, the effect of task instructions seems to diminish for difficult tasks compared to easier ones (Parker & Barber, 1964). In this study, participants who received task motivating instructions performed significantly better on a simple digit task than those who did not receive such instructions. However, for more difficult memory and reasoning tasks, no such differences in performance were found. The authors concluded that difficult tasks are not easily influenced by instructions. Finally, in an article by Koch (2008), an investigation of the role of task instructions in task repetitions and switches is reported, where it was found that the task instruction effect was present in the former and not the latter.

The current article reports an experiment where the effect of task instruction on DRT RT performance was assessed. Whereas most previous research investigated the effect of task instruction on the performance of one or two tasks, this experiment required participants to perform the DRT in addition to two other tasks; namely, a simulated driving task and either a cognitive task (n-back) or a primarily visual-manual task (Surrogate Reference Task; SuRT). As the DRT is of particular interest in in-vehicle testing, the simulated driving task was included in the study. The secondary tasks were chosen based on their relevance in the current DRT standardization process; precisely because of this, experience with these tasks and their effects according to the DRT has already been established.

Methods

Design

This experiment used a 4 (task set instruction) x 6 (task set load) mixed-model design. The between-subjects variable was task set instruction and the within-subjects variable was task set. Four participant groups were differently instructed on how to perform the same task sets. Participants performed the head-mounted version of the DRT (HDRT) in addition to two other tasks. The HDRT was tested under 6 types of task set load, including 2 baselines: HDRT performed alone and HDRT performed with the simulated driving task; and 4 experimental trials with 2 secondary tasks, the n-back and SuRT, in 2 levels of difficulty: HDRT + simulated driving task + n-back (0), HDRT + simulated driving task + n-back (2), HDRT + simulated driving task + SuRT easy, and HDRT + simulated driving task + SuRT hard.

Participants

Participants were separated into four groups reflecting the number of task set instructions tested. A total of 79 participants were tested, 71 of which qualified for analysis (2 were excluded due to ambiguous license status; 3 due to red-green colour blindness; 3 because of a technical error occurring in the data recording phase). All participants held a valid driver’s license and reported to have normal or corrected-to-normal vision. Table 1 describes these participants in more detail.

Table 1. Participant group demographics (N = 71).

Participant Group	*N*	*Mean Age*	*Minimum Age (years)*	*Maximum Age (years)*	*n Male*	*n Female*	*Left Handed*	*Right Handed*
1	18	24	17	33	16	2	2	16
2	18	24	19	33	15	3	1	17
3	18	24	19	28	13	5	3	15
4	17	23	18	29	12	5	4	13

Apparatus

This experiment was conducted in a fixed simulator at the Institute of Ergonomics, Technische Universität München. The driver’s seat was centrally located behind an active steering wheel (reconfigurable active yolk from Wittenstein). A large LCD monitor displayed the simulated driving scene (SILAB; Veitshöchheim, Germany), in front of the driver. To the right of the driver, a separate screen displayed the visual-manual task (Surrogate Reference Task, SuRT).

Head-Mounted Detection Response Task (HDRT)
The DRT used in this experiment is a USB device developed at the Institute of Ergonomics. The HDRT was implemented. The HDRT is a construction with one red LED mounted to a baseball-type cap and placed on the participant's head. The LED was viewed at 18 cm (measured along the centre line of the hat's brim, starting from where the brim is joined to the cap). Signals (viz. LED turning on) were randomly presented every 4000 – 5000 ms (signal onset to onset) and remained on for 1000 ms or until button press. The response button was fastened to the left index finger with a Velcro strip and positioned to facilitate a response by pressing the button against the steering wheel. Participants were instructed to respond as quickly and accurately as possible to these signals via button press.

Simulated driving task
The driving-like task simulated a two-way highway with two lanes in each direction. This simulated driving task required participants to drive along the highway scene in the middle of the right-hand lane, maintaining a constant speed of 80 km/h. The vehicle was automatic and required no gear shifting.

Secondary Tasks: n-back and SuRT
The n-back and the SuRT were used in both easy and difficult variants (n levels 0 & 2; Mehler et al., 2009; and SuRT as per ISO/TS 14198:2012) to induce different levels of workload. The n-back task is a system-paced task where pre-recorded numbers are dictated to the participant and he or she is to repeat these numbers. The "n" indicates the number the participant needs to repeat: either the number just said, n = 0, or the number said two steps prior to the current number, n = 2. In this task, participants were dictated 20 numbers and responded according to the current "n" of the task. Participants were to be as accurate as possible and responses were given verbally and recorded. The SuRT task required participants to find a pop-out target circle (larger circle) in a display filled with distracters (smaller, same sized circles). Task difficulty increased as the difference between the target and distracters was less prominent. Participants were to perform this task as quickly and accurately as possible. Navigation through and responses to the SuRT were performed via a numeric keypad located under the screen on the participant's right side.

Task set instructions
In the current experiment, a distinction between task instruction and task set instruction is made: task instructions indicated how to perform each individual task (given to all participants); task set instructions specified how to prioritize the performance of each task within the task set (given to all but one group, group 4 in Table 2). Each participant group (4 groups in total) was given different task set instructions with the goal of explicitly directing participants' attentive and cognitive resources. Table 2 presents the priorities specified by the task set instructions per participant group. These instructions were derived based on previous DRT literature and reasonable permutation. For safety reasons, the simulated driving task was always the primary priority for task set instructed groups.

Table 2. Instructed task priority per participant group.

Participant Group	*First Priority*	*Second Priority*	*Third Priority*
1	Simulated driving task	DRT	Secondary task
2		No priorities specified	
3		Secondary task	DRT
4	No priorities specified		

Procedure

At the beginning of the experimental session, participants watched a standardized multimedia instructional presentation that introduced and explained how to perform each task (viz. HDRT, simulated driving task, n-back, and SuRT) and how to prioritize the whole task set (ex. Driving + HDRT + either n-back or SuRT). Participants were able to ask questions and practice the tasks until they felt comfortable. Participants then performed the experiment, beginning with either a baseline or experimental block. A synchronization program ran all task software. Each trial began with allowing the participant to adjust their lane position and speed to 80 km/h. After performance on the simulated driving task was stable, the HDRT and secondary task began to run, respectively, within seconds of each other. Trials were recorded for 1 minute and included the simultaneous performance of all three tasks. Performance on the simulated driving-task and secondary tasks were also recorded, however, as the focus of the current article is on the DRT RT, other task performances are not discussed.

Dependent variables and hypotheses

HDRT RTs (ms) to HDRT signals were measured. Signal responses were classified as a hit between 100 ms and 2000 ms post-signal onset. Only the HDRT RTs of hits qualified for analysis. The RT value for a task set or baseline was calculated as the mean HDRT RT during this 1 minute segment. Means and standard deviations were then calculated across participants per task or baseline. The authors hypothesized that both task set load and task set instructions would both affect mean HDRT RTs. It was expected that task set instructions prioritizing the HDRT would yield faster mean RTs than those instructions where the HDRT was given a lower priority. For groups where task priority was less restricted (viz. group 2 and 4), similar DRT performance for these groups was expected due to a possible ingrained ascription of importance for the driving task. Additionally, it was hypothesized that task set would affect mean HDRT RTs, showing the sensitivity of the HDRT to different task loads.

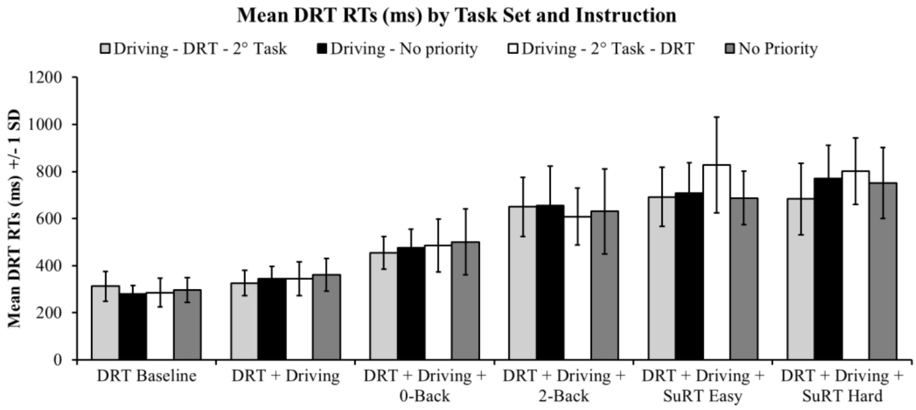

Figure 1. Mean HDRT RT (ms) +/- 1 SD per task set per task set instruction group. "DRT Baseline" indicates the performance of the HDRT as a single task.

Results

In Figure 1, the mean HDRT RTs (ms) across conditions per task set and instruction can be found. For all conditions, the hit rates were above 70%. It can be seen that in the baseline trials, mean HDRT RTs were overall faster than in the experimental trials. HDRT RT values, means (M) and standard deviations (SD), can be found in Table 3.

Data was analysed using a 4 (task set instruction) x 6 (task set load) mixed design ANOVA. Mauchly´s test indicated that the assumption of sphericity had been violated for task set load, $\chi 2$ (14) = 131.95, $p < .01$. Therefore the Greenhouse-Geisser correction ($\varepsilon = .68$) was used. No significant main effect of task set instruction was found on mean HDRT RTs.

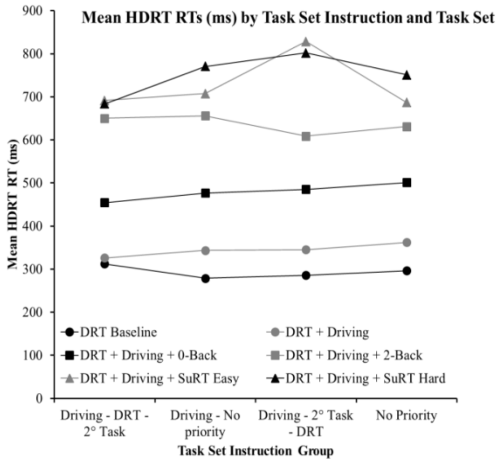

Figure 2. HDRT performance changes across task set instruction. Mean HDRT RT (ms) in each task set, except for the SuRT, was similar across task set instruction groups.

A significant interaction between task set load and task set instruction was found, $F(10.23, 228.46) = 2.74$, $p < .01$, $\eta_p^2 = .11$, meaning that depending on the task set load, the task set instruction had a different effect on mean HDRT RTs (see Figure 2). Simple effect tests, using Bonferroni adjusted alpha levels, indicated that for participant group 1, mean HDRT RTs did not significantly differ between the two baselines, DRT baseline and DRT + driving. Additionally, in groups 1, 2, and 4, mean HDRT RTs were not significantly different across task sets with 2-Back and SuRT in both difficulty levels. In group 3, no significant difference was found in sets with SuRT, easy and hard. All other comparisons were significant, $p < .01$. Pairwise comparisons were also carried out for each task set across participant groups. Task sets always yielded similar HDRT performance despite task instruction except for the task set with SuRT easy, under which mean HDRT RTs were significantly different between participant group 1 and 3, and 3 and 4, $p < .05$.

In terms of main effects, task set load was found as significant, $F(3.4, 228.46) = 312.43$, $p < .01$, $\eta_p^2 = .83$, meaning that the HDRT was sensitive to load induced by the different task sets. Specifically, the Bonferroni post hoc test revealed significant differences between HDRT RTs in all conditions, $p < .01$, except for the two task sets containing SuRT Easy and SuRT Hard.

Table 3. For each participant group, the specific task instruction is indicated as well as the mean (M; upper value) HDRT RTs (ms) and SDs (lower value).

	Task Instruction	Baselines		DRT + Driving +			
		DRT Alone	DRT + Driving	n-Back (0)	n-Back (2)	SuRT Easy	SuRT Hard
1	Driving+DRT+2°Task	312.32	326.32	454.67	650.33	691.52	683.67
		62.74	53.88	68.86	125.50	125.46	151.63
2	Driving+No Priorities	278.80	343.77	476.67	655.90	707.66	770.80
		36.85	52.79	79.39	167.98	129.13	139.86
3	Driving+2°Task+DRT	285.85	345.08	485.08	608.77	828.21	801.84
		60.18	71.93	112.07	120.22	203.08	140.86
4	No Priorities	296.67	362.20	501.18	631.02	687.03	750.77
		53.35	69.19	139.84	180.36	113.30	150.94

Discussion

The aim of the current experiment was to investigate whether different task set instructions affected mean DRT RTs. The HDRT was used for this evaluation, as were a simulated driving task and two additional secondary tasks: n-back and SuRT, in two levels of difficulty. It was hypothesized that in addition to the task set load affecting the DRT metric, task set instructions prioritizing the HDRT would yield faster RTs from participants relative to those instructions where the HDRT had a lower priority. Additionally, when no task priority was instructed, similar DRT performance for these groups was expected. Although task set load did affect mean DRT RTs, no significant effect of task instruction was found on mean HDRT RTs.

Generally, the data trends indicate that for all task sets with lower load (i.e. easier task variants), except for the SuRT, faster HDRT RTs were yielded. These HDRT RTs increase with task demand. A significant interaction between task set instruction and task set was found and revealed that mean HDRT RTs differed according to both task set load and instruction. By visually inspecting the graphs in Figure 1 and 2, it can be seen that no systematic influence seems to affect the data. The source of this interaction stems from the SuRT task. According to the HDRT RTs, SuRT easy did not always lead to faster RTs than for SuRT hard, in fact, sometimes this relationship was reversed. Since the DRT method proposes to measure the attentional effects of cognitive workload, it is not surprising that such effects are found for a primarily visual-manual task, which differs more in visual demand than it does in cognitive (Young et al., 2013). Additionally, because the SuRT is self-paced, participants could have used a different performance strategy relative to other tasks, which were system paced.

The significant main effect of task set load shows that the HDRT was sensitive to the changes in load corresponding with each task set. In accordance with this finding, it can be assumed that the participants were, in fact, performing the tasks in all conditions, else, no significant difference would be expected. As discussed above, the differences in HDRT performance during the SuRT easy and difficult was not significant.

The main effect of task instruction was not significant, implying that the task instructions used in this experiment did not affect mean HDRT RTs. At this time it cannot be confirmed whether task instruction, therefore, has no effect on the performance of a task set involving multiple tasks. First, an in depth evaluation of the other performance metrics, including an analysis of the performance strategies used in solving the tasks, needs to be carried out. It is possible that the difference in task performance manifests in another performance metric. The evaluation of the additional task performances included in this study is currently on-going.

In terms of the current experiment and the manipulation used, the mean HDRT RTs were not significantly affected by altering task instructions. Alternative hypotheses for this occurrence could be that because participants had to perform multiple tasks at once, the task set was too complex and participants were not able to accurately attend and perform as instructed (supported by Parker et al., 1964). As there were many tasks included within this experiment, it is also possible that some sort of task switching effect affected the performance metric, similar to the finding described by Koch (2008). The exact dynamic present here needs to be further investigated. Future investigations should also examine if the way the DRT and other tasks were technically operated (see Methods, each trial began with the driving simulation, then the DRT and secondary task started within seconds of each other) affects the performance outcome. This could indicate that giving different task instructions is not enough to alter task performance (supported by Levy et al., 2001) and that participants were not able to or did not adapt their strategy to the instructions given. An additional subjective questionnaire on how participants felt they were able to follow the instructions given could help clarify any inexplicable trends in the data. Future reproductions of this study could also include additional incentives for participants to ensure that the tasks are performed as close to the instructed priority as possible.

Conclusion

In terms of the current experiment, it has been demonstrated that task set instruction does not significantly affect mean HDRT RTs. Further evaluation is required to definitively confirm that task set instructions, therefore, have no measureable effect on task performance in a multiple task setting.

Acknowledgements

The authors would like to thank B. Helfer for his assistance in collecting the data, and A. Eichinger and M. Körber for their experimental and statistical insight.

References

Ansorge, U., & Neumann, O. (2005). Intentions Determine the Effect of Invisible Metacontrast-Masked Primes: Evidence for Top-Down Contingencies in a Peripheral Cuing Task. *Journal of experimental psychology. Human perception and performance*, *31*, 762–77.

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.

Bruyas, M.P., & Dumont, L. (2013). Sensitivity of Detection Response Task (DRT) to the driving demand and task difficulty. In *Proceedings of the Seventh International Driving Symposium on Human Factors in Driver Assessment, Training, and Vehicle Design* (pp. 64–70). Iowa City, USA: University of Iowa.

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

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

De Waard, D. (1996). *The Measurement of Drivers' Mental Workload*. (Doctoral Thesis). University of Groningen. Haren, The Netherlands: University of Groningen, Traffic Research Centre.

Engström, J., Åberg, N., Johansson, E., & Hammarbäck, J. (2005). Comparison between visual and tactile signal detection tasks applied to the safety assessment of in-vehicle information systems. In *Third International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design* (pp. 232–239). Iowa City, USA: University of Iowa.

Engström, J., Larsson, P., & Larsson, C. (2013). Comparison of static and driving simulator venues for the tactile detection response task. In *Proceedings of the Seventh International Driving Symposium on Human Factors in Driver Assessment, Training, and Vehicle Design* (pp. 369–375). Iowa City, USA: University of Iowa.

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*, 145–158.

Haider, H., & Frensch, P.A. (1999). Information reduction during skill acquisition: The influence of task instruction. *Journal of Experimental Psychology: Applied*, *5*, 129–151.

Harbluk, J.L., Burns, P.C., Hernandez, S., Tam, J. & Glazduri, V. (2013). Detection response tasks: Using remote, headmounted and Tactile signals to assess cognitive demand while driving. In *Proceedings of the Seventh International Driving Symposium on Human Factors in Driver Assessment, Training, and Vehicle Design* (pp. 78–84). Iowa City, USA: University of Iowa.

ISO/CD 17488: Road vehicles - Transport information and control systems – Detection-Response Task (DRT) for assessing selective attention in driving, 2013-09, International Organization for Standardization, Geneva, Switzerland. http://www.iso.org/iso/catalogue_detail.htm?csnumber=59887

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

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*, 255–275.

Knowles, W.B. (1963). Operator Loading Tasks. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *5*, 155–161.

Koch, I. (2008). Instruction effects in task switching. *Psychonomic Bulletin & Review*, *15*, 448–452.

Levy, J., & Pashler, H. (2001). Is dual-task slowing instruction dependent? *Journal of Experimental Psychology: Human Perception and Performance*, *27*, 862–869.

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.

Ogden, G.D., Levine, J.M., & Eisner, E.J. (1979). Measurement of Workload by Secondary Tasks. *Human Factors : The Journal of the Human Factors and Ergonomics Society*, *21*, 529–548.

Olsson, S., & Burns, P. C. (2000). Measuring driver visual distraction with a peripheral detection task. Retrieved December 13, 2013, from http://www-nrd.nhtsa.dot.gov/departments/Human Factors/driver-distraction/PDF/6.PDF

Trottier, L., & Pratt, J. (2005). Visual processing of targets can reduce saccadic latencies. *Vision research*, *45*, 1349–54.

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* (TNO-Report No. TM-99-C043). Soesterberg, Netherlands: TNO Human Factors.

Wenke, D., & Frensch, P.A. (2005). The Influence of Task Instruction on Action Coding: Constraint Setting or Direct Coding? *Journal of experimental psychology. Human perception and performance*, *31*, 803–19.

Wickens, C.D., & Hollands, J.G. (2000). Attention, Time-Sharing, and Workload. In N. Roberts & B. Webber (Eds.), *Engineering Psychology and Human Performance* (Third Edit., pp. 439–479). Upper Saddle River, New Jersey: Prentice-Hall Inc.

Young, R.A., Hsieh, L., & Seaman, S. (2013). The Tactile Detection Response Task: Preliminary validation for measuring the attentional effects of cognitive load. In *Proceedings of the Seventh International Driving Symposium on Human Factors in Driver Assessment, Training, and Vehicle Design* (pp. 71–77). Iowa City, USA: University of Iowa.