

DETECTION RESPONSE TASKS: USING REMOTE, HEADMOUNTED AND TACTILE SIGNALS TO ASSESS COGNITIVE DEMAND WHILE DRIVING

Joanne L. Harbluk, Peter C. Burns, Sebastian Hernandez, Jane Tam & Viliam Glazduri
Transport Canada
Ottawa, Ontario, Canada
Email: joanne.harbluk@tc.gc.ca

Summary: Three versions of the Detection Response Task (DRT) were evaluated and compared as methods to assess cognitive demand while driving. Participants performed tasks known to manipulate cognitive demand (N_Back tasks) and real-world cognitive demand tasks using the iPhone interface, Siri, while driving a simulator. The Remote DRT, the Head Mounted DRT and the Tactile DRT were all sensitive to task demands but advantages were found to using the newer versions for the assessment of cognitive distraction. This work is part of an international collaborative research effort to standardize the DRT (ISO WD17488).

INTRODUCTION

Considerable effort has been directed toward developing effective methods to quantify distraction arising from interactions with other devices while driving. The Japan Automobile Manufacturers Association Guideline (JAMA, 2004), the European Statement of Principles on the Design of Human Machine Interaction (ESoP, 2005) and the Alliance guidelines (Alliance, 2006) were all developed to limit driver distraction from in-vehicle information and communication systems. These approaches provided a variety of design guidelines and test procedures to assess the performance of drivers interacting with other devices while driving. All these efforts have attempted to address distraction arising from interactions with visual manual interfaces. Most recently, the National Highway Traffic Safety Administration (NHTSA, 2012) also issued guidelines focused on visual-manual distraction from certain secondary, non-driving related original equipment tasks. While driver distraction from visual-manual tasks is certainly a priority, there are other types of distraction that remain to be addressed.

As in-vehicle interfaces evolve, voice user interfaces are becoming a common mode of interaction for drivers. It has been known for some time that the use of voice interfaces can negatively impact safe driving performance (e.g., Lee et al, 2001). Consequently, it is recognized that interference from poorly designed voice interfaces must be limited and that test methods are needed for these assessments. As part of its Driver Distraction Program (2010), NHTSA intends to extend the development of their guidelines in subsequent phases to cover voice interfaces.

In the present work, the Detection Response Task (DRT) was used to assess cognitive demand in the context of driving. A precursor of the DRT (the PDT, or peripheral detection task) showed promise in earlier work as a method to assess distraction arising from interactions with in-vehicle systems requiring speech commands (Ranney et al., 2005) and where cognitive demand was elevated due to information processing demands of the content and quality of speech output (Harbluk et al., 2005). These detection tasks have evolved from the original version which was

usually implemented as lights to be detected in the driving scene (e.g., van Winsum et al., 1999; Olsson & Burns, 2000; Jahn et al., 2005) to include a head mounted version (van der Horst et al., 2010), and a tactile version as well (Engström et al., 2005; Merat & Jamson, 2008).

A multisite International Standards Organization research project was undertaken to investigate the sensitivity, validity and effectiveness of the DRT to discriminate among different levels of cognitive demand while driving (ISO 17488, 2012). This collaborative project incorporated a number of tasks and manipulations with some studies conducted in simulators and others on road. Here, we focus on a simulator study and the assessment of cognitive demand experienced by the drivers as determined by three versions of the DRT, the Remote, the Head mounted and the Tactile. Two questions guided this research: 1) Are the three versions of the DRT differentially sensitive to increased levels of cognitive demand? 2) Are particular versions of the DRT better suited to detecting cognitive distraction of the sort that would arise from the use of speech based interfaces in the driving environment? The approach was to have drivers perform an artificial task known to impose increasing levels of cognitive demand (the N_Back task; Mehler et al., 2012; Ranney et al., 2011), to examine how the increased cognitive demand would be reflected in performance on each of the DRT versions. In addition, two real world tasks, where participants interacted with Siri, the voice interface for the iPhone, were included.

METHOD

Participants & Design

Sixteen participants (8 female, 8 male) aged 21 to 46 (mean = 29.4 years) volunteered to take part in the study. All were licensed drivers and estimated that they drove at least 10,000 km a year. A complete within-subjects design was used. The order in which participants performed the three DRT versions was counterbalanced using a balanced Latin square. The order of the tasks was set up such that all participants first drove a baseline section, then completed drives with the secondary tasks and an additional baseline section. Secondary tasks included the N_Back task with two levels of difficulty and two tasks involving interactions with Siri using the iPhone. The conditions for the secondary tasks (N_Back and Siri) were controlled such that the presentation order of the two types of task was counterbalanced across participants and the order of the two versions within the tasks was alternated across participants. Hit Rate and Reaction Time were collected for the DRT versions and workload ratings (1=low; 10=high) were collected for each task/DRT combination.

Equipment & Materials

Participants drove the National Advanced Driving Simulator (NADS) “MiniSim”. Stimulus presentation for the DRT was controlled by the simulator software and the DRT software provided by TNO. For the remote DRT, the stimulus was placed in the driving scene, at a distance of 0.98 m from the driver. The stimulus was a 5 mm red circle, placed at a vertical angle of 2 degrees above the horizon and at a horizontal angle of 11 degrees to the left of the middle of the screen. In the tactile detection task, the stimulus (tactor) was a 3.68 cm² vibrotactile pad taped to the shoulder. In the head mounted version, participants wore a headset that consisted of a 5 mm diameter LED on the end of a 21 cm extension. For each DRT version, the stimulus

would appear at an interval of 3-5 seconds apart, and would last 1 second if the participant did not respond to it. Responses to stimuli were collected up to 2 seconds after their appearance. (See Figure 1).

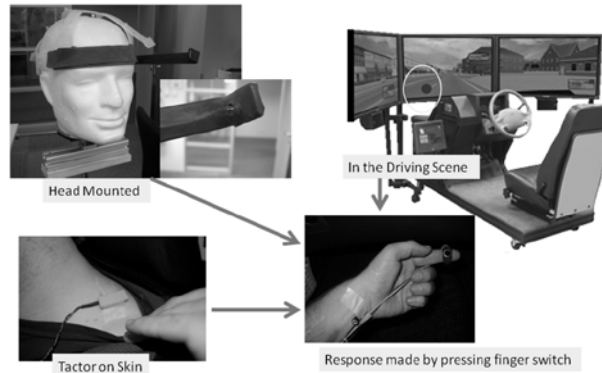


Figure 1. Set Up for the Remote, Head Mounted and Tactile Detection Response Tasks

Procedure

Participants were given an overview of the purpose and procedure of the study after which the consent form and demographics questionnaire were completed. Participants practiced the secondary tasks individually. The auditory stimuli for the N_Back consisted of a string of recorded single digits. In the easier version of the N_Back, the N_Back 0, the participant was required to simply repeat each number aloud after they heard it. In the more difficult version, the N_Back 1, participants were instructed to repeat the number that preceded the last one they heard (see Table 1).

Table 1. Examples of stimuli and responses for the N_Back task

N	I say	three	two	six	seven	one
N_Back 0	You say	<i>three</i>	<i>two</i>	<i>six</i>	<i>seven</i>	<i>one</i>
N_Back 1	You say	-	<i>three</i>	<i>two</i>	<i>six</i>	<i>seven</i>

Two tasks required interaction with iPhone’s Siri speech interface. In the SIRI_1 task, participants asked Siri questions from a predetermined list. In the other, SIRI_2 task, participants asked Siri to read a text message consisting of a message regarding two appointment dates. Participants were instructed to remember those dates and to instruct Siri to schedule these appointments in the iPhone’s calendar. Examples of the materials for the Siri interactions are presented in Table 2.

Participants also practiced driving only the simulator for three minutes. They were reminded to try to keep their speed at 100 km/hour and to stay in the right-most lane throughout the entire drive. The roadway consisted of a moderately curved 2 lane freeway with some ambient traffic. Practice was also provided for combinations of the tasks and DRT versions while driving. After practice, the experiment began. Participants completed the secondary tasks for one minute each while driving and performing one of the versions of the DRT. In the 30s break between tasks, participants provided their Workload Rating for the task they just completed using a scale of 1=low to 10=high.

Table 2. Examples of Materials used for the Siri Tasks

SIRI_1	SIRI_2
Questions posed to Siri Participant asks: What time is it? What is today's date? What's the temperature outside? Will it rain today? Will it be sunny on Friday? What's on my calendar for Friday? What day of the week is the 20 th ? Set an alarm for 8 am. What date is it on Saturday? Turn off the 8 am alarm.	Read Text & Make Calendar Appointment Participant says: "Read me my text" Siri says: "Your dentist appointment is Monday at 3. Your doctor appointment is Wednesday at 11. Make 2 entries in your calendar." Participant says: "Schedule appointment Monday at 3 Schedule appointment Wednesday at 11."

RESULTS

Data were analyzed using repeated measures ANOVA and LSD for tests of means.

Hit Rate

Hit Rates exceeded 80% in all conditions, surpassing the 70% requirement of the ISO procedure.

Reaction Time

N_Back Tasks. Mean reaction times for Baseline and the N_Back tasks are displayed for the three DRT versions in Figure 3. The reaction times increased from Baseline to N_Back 0 to N_Back 1 for all three DRT versions. It was only for the Tactile DRT, however, that significantly longer reaction times were exhibited as task difficulty increased from Baseline, to the easier and then the more difficult level of N-back task ($F(2,30)= 18.42, p<.0001$; all pairwise means $p<.01$).

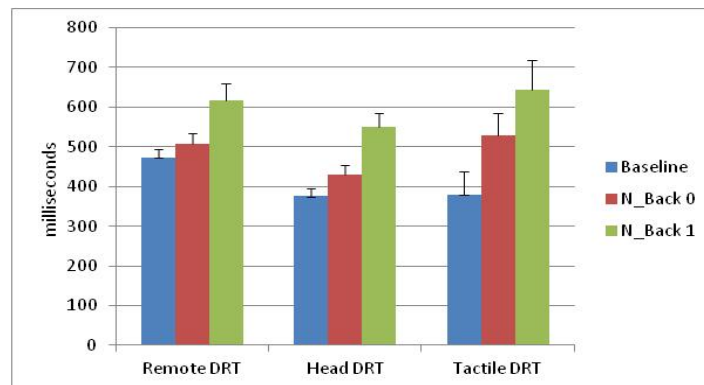


Figure 3. Mean Reaction Times (+SE) for the N_Back Tasks for Each Version of DRT

For the Remote DRT, mean reaction times were significantly longer for the N_Back 1 compared with the N_Back 0 and Baseline ($F(2,20)= 17.29, p<.0001$). Results for the Head mounted DRT followed the same pattern with N_Back 1 reaction times longer than both the Baseline and the N_Back 0 conditions $F(2, 30)= 16.42, p<.0001$. No significant differences were found between the Baseline and N_Back 0 condition for the Remote or Head mounted DRTs. Participants' ratings of workload increased significantly across the three conditions of No Task, N_Back 0 and N_Back1 for all three versions of the DRT.

Siri Tasks. Figure 4 displays reaction times obtained using the DRT versions when Siri tasks were undertaken while driving. Participants took longer to respond to the stimuli when engaged in the Siri tasks than during Baseline performance, indicating that the DRT tasks were sensitive to the cognitive demands imposed by Siri interaction. This was observed for the Remote DRT ($F(2,30)= 21.88, p<.0001$), the Head Mounted DRT ($F(2,30)= 20.44, p<.001$) and the Tactile

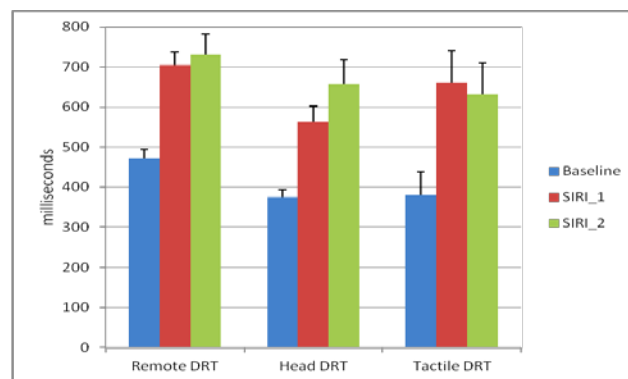


Figure 4. Mean Reaction Times (+SE) for the Siri Tasks for Each Version of DRT

DRT ($F(2,30)= 20.82, p<.0001$). In all cases reaction times were longer for each of the Siri tasks individually compared with Baseline ($p<.001$). Interestingly, the email/calendar version (SIRI_2) resulted in longer times than the multiple question SIRI_1 task, but only in the Head Mounted DRT version ($p<.05$). Participants' workload ratings indicated that they experienced the two Siri tasks as more demanding than Baseline for all three versions of DRT. No differences were observed in workload ratings between Siri tasks except in the Tactile DRT condition, where the Siri email/calendar task was rated as more demanding than the multiple question task.

DISCUSSION

As voice user interfaces become more common in vehicles, methods are needed to assess the cognitive demands associated with their use. A method is desired that is sensitive, objective, reliable, valid & practical (Burns et al., 2010). The first question to be addressed in the current research concerned the sensitivity of the DRT to levels of cognitive demand. The results from the N_Back task, which provided a predetermined manipulation of cognitive load, indicated that all three DRT versions distinguished the more demanding N_Back task from the easier one. Only the Tactile DRT, however, was sensitive to increases in cognitive demand from Baseline as well. The utility of the DRT to assess cognitive demand from real world tasks was assessed using Siri, the speech interface for the iPhone. The additional cognitive demand of interacting with Siri was reflected in longer reaction times compared with Baselines for all three DRT versions.

Interestingly, longer times for the calendar task compared with the question task were found when the Head mounted DRT was used. Specific manipulations of cognitive demand in real world tasks will need to be explored in future work.

There are other issues to be considered. Task testing duration will need to be addressed. In the current work, all tasks were carried out for one minute durations. Should the testing occur for a predetermined duration or for the duration of the task as it naturally occurs? The method will have to be assessed for suitability for short real world tasks (e.g., < 10 seconds) with consideration of issues such as the minimum number of stimuli that must be presented for reliable measures to be obtained. It is outside the scope of the ISO work, but performance criteria will need to be established for the use of the DRT as a method to assess cognitive distraction and decide which tasks are compatible or not with safe driving performance.

The second question had to do with whether particular versions of the DRT are better suited to detect cognitive distraction of the sort that would arise from the use of speech based interfaces in the driving environment. Although all three DRT versions were sensitive to cognitive demand there are additional considerations. The Remote DRT can be more difficult to implement consistently than either the Head mounted or Tactile versions in simulators or on road. One important limitation for the Remote DRT is that, while it may be sensitive to cognitive distraction, it is also affected by visual distraction. When the driver's eyes are directed away from the driving scene they can miss the stimuli presented ahead. Both the Head mounted and Tactile versions overcome this confound since the stimuli are always "with" the participant. Finally, an advantage of the Tactile DRT is that it can be used without interfering with eye tracking equipment should it be used.

The newer variants of the DRT appear to be useful tools to assess cognitive distraction. The results herein, and from the other ISO studies, will add to our understanding of its utility.

REFERENCES

- Alliance of Automobile Manufacturers Driver Focus-Telematics Working Group, "Statement of Principles, Criteria and Verification Procedures on Driver-Interactions with Advanced In-Vehicle Information and Communication Systems, June 26, 2006 Version.
- Burns, P.C., Bengler, K., & Weir, D.H. (2010). Driver Metrics, an Overview of User Needs and Uses. In G. Rupp (Ed.) *Performance Metrics for Assessing Driver Distraction: The Quest for Improved Road Safety*. Warrendale, PA: SAE International.
- Engström, J. (2010). The Tactile Detection Task as a Method for Assessing Drivers' Cognitive Load. In G. Rupp (Ed.) *Performance Metrics for Assessing Driver Distraction: The Quest for Improved Road Safety*. Warrendale, PA: 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 Information Systems," *Proceedings of the Third International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design*. Rockport, Maine, 232-240.
- ESOP, European Statement of Principles on the Design of Human Machine Interaction (ESoP 2005): Draft, European Commission Information Society and Media Directorate-General - G4 ICT for Transport, 2005.

- Harbluk, J.L. & Lalande, S. (2005). Performing Email Tasks While Driving: The Impact of Speech-Based Tasks on Visual Detection. *Proceedings of the Third International Driving Symposium on Human Factors in Driving Assessment, Training, and Vehicle Design*. Rockport, Maine, 311-317.
- ISO 17488 (2012). Road Vehicles- Transport information and control systems- Detection Response Task. (New Work Item for ISO TC 22 SC13 WG8).
- 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(3), 255-275.
- JAMA Japan Automobile Manufacturers Association Guideline for In-vehicle Display Systems (JAMA), 2004. http://www.jama-english.jp/release/release/2005/In-vehicle_Display_GuidelineVer3.pdf
- 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, 631-640.
- Mehler, B., Reimer, B. & Coughlin, J.F. (2012). Sensitivity of Physiological Measures for Detecting Systematic Variations in Cognitive Demand from a Working Memory Task: An On-road Study Across Three Age Groups. *Human Factors*, 54(3), 396-412.
- Merat, N. & Jamson, H.A. (2008). The Effect of Stimulus Modality on Signal Detection: Implications for Assessing the Safety of In-Vehicle Technology, *Human Factors*, 50, 145-158.
- NHTSA (2010). *Overview of the National Highway Traffic Safety Administration's Driver Distraction Program* NHTSA Distracted Driving Research Plan - April 2010 [PDF] DOT-HS-811-299
- NHTSA (2012). Visual-Manual NHTSA Driver Distraction Guidelines for In-Vehicle Electronic Devices <http://www.gpo.gov/fdsys/pkg/FR-2012-02-24/pdf/2012-4017.pdf>.
- Olsson, S., & Burns, P. C. (2000). *Measuring Driver Visual Distraction with a Peripheral Detection Task*. Paper presented at the NHTSA Internet Distraction Forum.
- Ranney, T.A., Baldwin, G.H.S., Parmer, E., Domeyer, J., Martin, J. & Mazzae, E.N. (2011). *Developing A Test To Measure Distraction Potential Of In-Vehicle Information System Tasks In Production Vehicles* (Rep. No. Dot HS 811 463). Washington, D.C.: U.S. Department Of Transportation, National Highway Traffic Safety Administration.
- Ranney, T.A., Harbluk, J.L. & Noy, Y.I. (2005). Effects Of Voice Technology On Test Track Driving Performance: Implications For Driver Distraction. *Human Factors*, 47, 439-454.
- 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. Rupp (Ed.) *Performance Metrics for Assessing Driver Distraction: The Quest for Improved Road Safety*. Warrendale, PA: SAE International.
- 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.