COMPARISON OF STATIC AND DRIVING SIMULATOR VENUES FOR THE TACTILE DETECTION RESPONSE TASK

Johan Engström, Pontus Larsson & Christian Larsson
Advanced Technology and Research, Volvo Group Trucks Technology
Gothenburg, Sweden
Email: johan.a.engstrom@volvo.com

Summary: The general objective of the present study was to validate a low-cost, static, version of the Tactile Detection Response Task (TDRT) intended for driver-vehicle interface evaluation in industrial settings. The static TDRT venue was compared to the more commonly used driving simulator venue, where the TDRT and the secondary task under evaluation are performed during simulated driving. The results indicated that the effect of venue was additive over a range of visual-manual and cognitive secondary tasks, which offers preliminary support for the static TDRT venue as a surrogate for the driving simulator TDRT venue. However, a more detailed analysis revealed a counterintuitive effect for one of the visual-manual secondary tasks (SuRT), where the easier version of the task (as confirmed by subjective workload ratings) yielded a stronger effect on the TDRT than the more difficult version. Possible explanations and implications for the TDRT and its application to driver-vehicle interface evaluation are discussed.

INTRODUCTION

During the past decade, the Detection Response Task (DRT; previously known as the Peripheral Detection Task, PDT; Olsson and Burns, 2000) has become a popular method for measuring the effects of driving and secondary task demand on driver attention, in particular in the context of driver-vehicle interface evaluation. The DRT measures such effects in terms of the performance on a simple detection-response task where the subject presses a button in response to stimuli presented at an interval of 3-5 seconds. The method is currently subject to ISO standardization (ISO, 2012).

One specific version of the DRT included in ISO (2012) is the Tactile DRT (TDRT), where the stimulus consists of a vibration issued by a small vibrator (tactor) attached to the body. The TDRT was first developed, and compared to the “classical” PDT, where stimuli are presented visually, in an on-road study by Engström et al. (2005). The TDRT has since then been further developed and demonstrated sensitive to both driving and secondary task load (Merat and Jamson, 2008; Mattes, Föhl and Schindhelm, 2008; Diels, 2011; see Engström, 2010, for a general review).

Traditionally, the DRT has been used in a triple task setting, where the DRT is performed concurrently with the secondary task under evaluation and driving, in a driving simulator or in real traffic. These are referred to as the *driving simulator* and *on-road venues* respectively (ISO, 2012). However, since driving simulator- and on-road studies are generally relatively costly and labor intensive, a simpler, *dual-task*, version of the DRT would be better suited for industrial application in product development. Such a dual-task setting, without a driving task, is referred to as the *static venue* in ISO (2012). However, a yet open issue is to what extent measurements obtained in the static venue can be generalized to the more ecologically valid driving simulator and on-road venues. Due to the lower overall level of task demand in the static venue (due to the absence of a driving task), it may be expected that absolute response performance is better in the static compared to driving simulator and on-road venues. However, as long as the effect of venue is *additive* across different secondary tasks (i.e., there is no interaction between venue and task), the static venue could, in principle, be used as a valid surrogate for the driving simulator venue.

The main objective of the present study was to examine this issue by comparing effects of different types of secondary task demand on TDRT performance in static and driving simulator venues. A key general issue is thus the extent to which the effect of venue interacts with the effects of secondary task demand. The presence of such an interaction would strongly question the validity of the static TDRT. A more specific objective was to investigate to what extent the two venues yield similar sensitivity to systematic manipulations of cognitive and visual-manual load. The study was part of a set of international coordinated studies, conducted under ISO TC22/SC13/WG8, with the general goal to support the ongoing DRT ISO standardization effort (ISO, 2012).

METHOD

Participants

16 male subjects participated in the study. Their age ranged from 29-51 years (mean 36.8, SD 6.6). All participants held a truck driver's license. They were given two cinema tickets for their participation.

TDRT implementation

The TDRT implementation generally adhered to the specifications in ISO (2012). The tactor was fixed to the participant's left shoulder by means of medical tape. The inter-stimulus interval was randomly uniformly distributed between 3 and 5 seconds. The stimulus was onset for maximum 1 s and extinguished when a response was given.

Venues

The fixed-base truck simulator at Volvo Advanced Technology and Research (ATR) was used as the driving simulator venue. The simulator consists of a full Volvo FH truck mockup positioned in front of a circular screen with a 135 degree horizontal field of view. The driving environment is displayed on the screen by means of three projectors with overlapping images. In the static venue, the participants were positioned in the driver's seat in the truck cabin and performed the secondary tasks without driving. Except for the lack of driving task, the static venue was thus identical to the driving simulator venue.

Simulated driving task

The driving task consisted of free driving on a motorway with moderate curvature. Some ambient traffic was present but did not directly interact with the participant's vehicle. The posted speed was 90 km/h and higher speeds were prevented by means of a speed limiter. Subjects were instructed to remain in the right lane throughout the drive.

Secondary tasks

Three secondary tasks were included, two artificial tasks (n-back and SuRT) which allowed a systematic manipulation of cognitive and visual-manual load respectively and one more naturalistic task (Siri). The n-back and SuRT difficulty was varied at two levels, thus yielding five secondary task conditions in total.

The *n-back* task (Mehler, Reimer and Dusek, 2011) required the driver to orally repeat numbers in a sequence of numbers read up by a recorded voice. In the 0-back condition, the participant just had to repeat the last number in the sequence. In the 1-back condition, the participant had to repeat the second but last number. Thus, 0-back and 1-back tasks only differed in terms of their demands on working memory (i.e., their cognitive load). The n-back task was implemented according to the specification in Mehler et al. (2011), with the voice instructions translated to Swedish.

In the *Surrogate Reference Task* (SuRT; specified in ISO, 2010; see also Mattes et al., 2007) subjects searched a visual display for a target circle which differed in size compared to a set of surrounding distractor circles. The visual perceptual load was varied at two difficulty levels in terms of the relative difference in size between the target and the distractor circles (smaller difference in the hard version). Once the target was found, the subjects responded by indicating its location on the display with a rectangular cursor, controlled by a keypad, that covered a portion of the display. The SuRT levels also differed in terms of in manual load, which was varied terms of the width of the cursor. For the SuRT Easy level, the cursor covered half of the display and thus had to be moved one step, or not at all, to reach the target. For the SuRT Hard level, the cursor was narrower and potentially had to be moved several steps to reach the target. The SuRT was paced by the subject. Thus, once a response was made, a new display appeared.

Finally, the *Siri* task involved interaction with the Siri function featured on the iPhone 4S. Siri is a voice-controlled personal assistant that provides answers to questions asked in natural language. Subjects were instructed to ask Siri pre-defined questions from a list posted inside the truck cabin. The iPhone was mounted on the dashboard within comfortable reach for the participant. Asking a question required the subject to push the “microphone” button on the iPhone touch screen. Visual feedback on the voice input, as well as Siri’s answer, was displayed on the screen.

All secondary tasks were initiated by a pre-recorded voice message and performed for one minute, until a voice message instructed the participant to stop. Baseline data (no secondary task) was also collected for one minute. The duration between task conditions was 30 seconds. Each task and baseline condition was repeated twice per venue.

Experimental design

The independent variables in the study were TDRT Venue {static, driving simulator} and secondary task {0-back, 1-back, SuRT Easy, SuRT Hard, Siri, Baseline}. Moreover, Task Repetition {first, second} was included as an independent variable to check for learning effects. These independent variables were all varied within-group. The order of the venue and secondary task conditions were counterbalanced between participants. As described below, the venue order determined the amount of secondary task training that the participant received prior to the static venue condition. Thus, Venue Order was included as a between-subject subject variable to check whether this difference in training had any effect on the results.

The main dependent variable was TDRT response time, but TDRT hit rate was analyzed as well. Based on the current version of ISO (2012), a hit was defined as a response made between 100 and 2500 ms after stimulus onset and the hit rate was calculated as the ratio between the number of hits and the total number of stimuli. Response time was defined as the time that lapsed between stimulus onset and response and was only computed for hits. In addition, secondary task performance and subjective workload was analyzed. Secondary task performance was measured in terms of percent correct responses/trials and only analyzed for the n-back and SuRT tasks. The workload was rated by the participant during the 30 s period between task conditions on a scale from 1 to 9, and reported verbally by the participant. In addition effects on various driving performance measures were analyzed but not reported here due to space constraints.

Procedure

Participants were first given a general introduction to the experiment and tasks. They then practiced each task (n-back, SuRT, Siri, TDRT) separately, guided by the experimenter. This was followed by practice on the dual task condition (TDRT + n-back/SuRT/Siri), guided by a pre-recorded voice. They then practiced driving in the simulator. Finally, the participants practiced the triple-task condition (driving + TDRT + n-back/SuRT/Siri). However, for the participants that performed the static venue first (half of all subjects), the experimental trials started directly after practicing the dual task condition. For this group, the practice on driving the simulated vehicle was conducted prior to the driving simulator trials. In total, the practice session took about 35-40 min to complete. Prior to the experimental trials, the participants were also given task priority instructions. Based on ISO (2012), participants were instructed to focus primarily on the driving task (in the driving simulator venue) and do their best their best to also perform the TDRT and the secondary task under evaluation.

RESULTS

The performance on the n-back and SuRT task was generally high. For SuRT, the mean percent correct responses exceeded 97% in all conditions. For the analysis of n-back performance, 29 of 128 data points were missing due to a technical error in the audio recording (data was missing completely for three participants and partly for two participants). Moreover, it was found that participants on four occasions performed the wrong n-back task (e.g., 0-back instead of 1-back and vice versa). For the data available, the mean percentage correct n-back responses exceeded 97% in all conditions (and 100% in both 0-back conditions).

For the DRT analysis, the four cases where the participant performed the wrong version of the n-back task were excluded from the analysis. In addition, for response time, one data point was missing due to the participant failing to produce any valid responses in that particular condition. In order to not lose these participants' data in the repeated measures analysis, these five missing data points were replaced by the data points obtained from the participants' second repetition of the same task for the same venue.

Hit rates for the TDRT were high (>84%) across secondary task conditions and approximated 100% for the baseline conditions in both venues. The data for mean response time across tasks for the two venues is plotted in Figure 1. A 4-way repeated measures ANOVA was conducted with Task (6 levels), Venue (2 levels), Task Repetition (2 levels) and Venue Order (2 levels, between groups). Due to a violation of the sphericity assumption, the degrees of freedom were corrected by Greenhouse-Geisser's method. There was a significant main effect of Task (F(2.9, 43.2)=43.6, p<.001) as well as a significant main effect of Venue (F(1, 15)=24.8, p<.001), but no main effect of Task Repetition or Venue Order. As expected, response times were consistently lower in the static compared to the driving simulator venue. Most importantly, however, there was no significant interaction between Task and Venue and no other significant interactions.

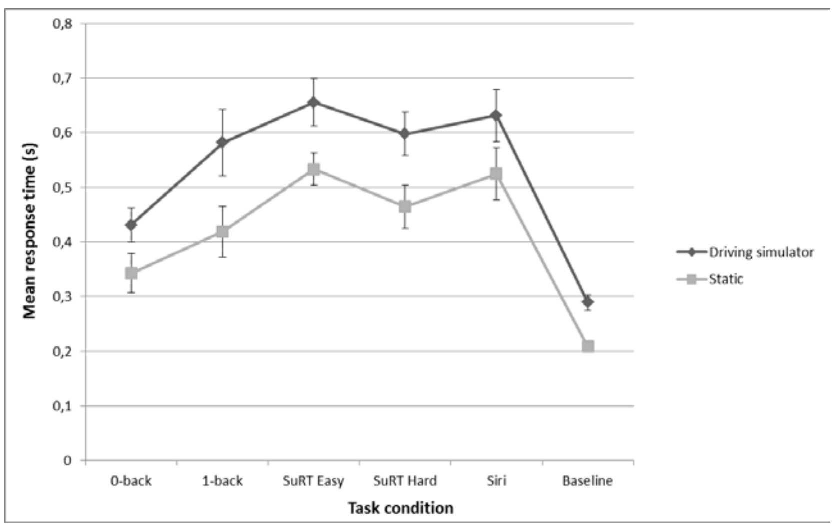

Figure 1. Mean response times acros the secondary task conditions for the driving simulator and static venues. The values for each task condition represent the average across all subjects in the two task repetitions. The error bars represent the standard error of the mean

In order to investigate the sensitivity of the two TDRT venues to systematic manipulations of cognitive and visual-manual load, separate ANOVAs along with planned post-hoc contrast analysis were conducted for the two venues. The contrasts of interest were those between 0-back vs. 1-back (cognitive load) and SuRT Easy vs. SuRT Hard (visual-manual load).

For the driving simulator venue, there was again a significant main effect of Task (F(3.1, 46.6)= 24.8, p<.001). There was also a significant main effect of Task Repetition (F(1, 15)=5.2, p<.05), indicating that some learning took place in this condition, but no effect of Venue Order. The 1-back condition yielded significantly longer mean response times than the 0-back condition (F(1, 15)=16.7, p<.01). However, there was no statistically significant difference between SuRT Easy and SuRT Hard.

In the static venue, there was also a significant main effect of Task (F(2.5, 37.3)=33.8, p<.001 but no significant effects of Task Repetition or Venue Order. The 1-back condition yielded significantly longer mean response times than the 0-back condition (F(1, 15)=10.0, p<.01). However, by contrast to the driving simulator venue, a significant difference was found between SuRT Easy and SuRT Hard (F(1, 15)=13.3, p<.01). Somewhat unexpectedly, as can be seen in Figure 1, the SuRT Easy condition yielded significantly *longer* mean response times than the SuRT Hard condition. In order to further explore this result, an analysis of subjective workload ratings was conducted for the SuRT Easy and SuRT Hard conditions. While a full exposition of these results is outside the scope of the present paper, the key finding was that workload ratings increased significantly with task difficulty, for both the n-back and the SuRT tasks, in both venues.

CONCLUSION

The main objective of the present analysis was to compare the effects of different types of secondary task demand on TDRT performance in static and driving simulator venues. In general, the additive effect of Venue across secondary tasks conditions (i.e., the lack of a venue-task interaction) supports the use of the static TDRT venue as a surrogate for more labor-intensive and costly driving simulator settings. However, it should be noted that this does not prove the null hypothesis (of no interaction). Furthermore, the driving task used in the driving simulator venue was relatively non-demanding, and the presently found additive effect of venue may not generalize to more demanding driving conditions where the total level of demand approaches the driver's capacity limits. Also, different types of secondary tasks may have different effects in different driving conditions. Hence, future studies, including a greater range of naturalistic secondary tasks and driving conditions, are needed to further test the validity of the static TDRT setup. It is also an open question to what extent the present results generalize to other versions of the DRT that present the stimulus visually rather than by tactile stimulation.

The more detailed analysis of contrasts between n-back and SuRT difficulty levels per venue revealed that, in the static venue, SuRT Easy yielded longer mean response times than SuRT Hard, which appears somewhat counterintuitive. The same tendency was also present in the driving simulator venue but did not reach statistical significance (see Figure 1). However, subjective workload ratings were significantly higher for SuRT Hard than for SuRT Easy in both venues. One possible explanation for this apparent discrepancy is that the SuRT Easy, due to its relative simplicity, was performed at a higher pace than SuRT Hard and thus involved more manual operations (button presses) per time unit (as mentioned above, the SuRT was self-paced). This was particularly pronounced in the static venue where the mean number of trials completed during the one-minute task period was 61.6 for SuRT Easy compared to 19.1 for SuRT Hard. The corresponding numbers for the driving simulator venue were 39.5 and 12.9 respectively. Thus, specific motor interference between TDRT responses (performed with the left hand) and SuRT button presses (performed with the right hand) was potentially stronger for SuRT Easy compared to SuRT Hard. This could explain the stronger effect of SuRT Easy on the TDRT, while participants still rated the SuRT Hard as more loading due to its higher perceptual demand. This explanation has the important implication that the DRT, especially when applied in the static venue, may not be the best choice for evaluating driver-vehicle interfaces with strong manual demands (e.g., requiring frequent button presses), since the resulting specific motor interfere may mask the more relevant effect of cognitive load. However, it should be noted that this effect has not been found in other, similar, studies. For example, Mattes et al., 2007, as well as several other (yet unpublished) ISO-coordinated studies, used the same SuRT levels and found no difference in TDRT response time. Thus the present results should be treated with caution.

In conclusion, the present results offer preliminary support for the use of a low-cost, static, TDRT venue for driver-vehicle interface evaluation in industrial settings, at least for tasks without excessive motor demands (such as e.g., voice control). A key benefit of the static TDRT venue is that it can be used in almost any setting (desktop, production vehicles etc.), making it ideal for application in product development. However, further work is needed to verify the present results for a greater range of secondary task and simulated driving conditions.

ACKNOWLEDGEMENTS

We would like to thank Prof. Richard Young, Wayne State University, for valuable comments and discussions during the drafting of this paper, and VINNOVA for partly funding this research.

REFERENCES

Engström, J. (2010). The Tactile Detection Task as a method for assessing drivers' cognitive load. In G. L. Rupp (Ed.), *Performance Methods for Assessing Driver Distraction: The Quest for Improved Road Safety* (pp. 90-103). Warrendale, PA: SAE International.

Engström, J., Åberg, N., Johansson, E. and Hammarbäck, J. (2005). Comparison between visual and tactile signal detection tasks applied to the safety assessment of in-vehicle information systems (IVIS). *Proceedings of the Third International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design*. Rockport, Maine.

Diels C. (2011). Tactile Detection Task as a real time cognitive workload measure. In M. Anderson (Ed.), *Contemporary Ergonomics and Human Factors 2011*, pp. 183-190, CRC Press.

ISO. (2010). Calibration tasks for methods which assess demand due to the use of in-vehicle systems. ISO PDTS14198, TC 22/SC 13/WG 08 (preparatory STD Version 2.2).

ISO (2012). Detection-Response Task for assessing selective attention in driving. New Work Item, ISO TC22/SC13/WG8.

Mattes, S., Föhl, U. and Schindhelm, R. (2007). *Empirical comparison of methods for off-line workload measurement*. AIDE Deliverable 2.2.7, EU project IST-1-507674-IP

Mehler, B., Reimer, B. & Dusek, J.A. (2011). *MIT AgeLab delayed digit recall task (nback).* MIT AgeLab White Paper Number 2011–3A. Massachusetts Institute of Technology, Cambridge, MA.

Merat, N. and Jamson, A.H. (2008). The effect of stimulus modality on signal detection: Implications for sssessing the safety of in-vehicle technology. *Human Factors*, Vol. 50, No. 1, pp. 145–158.

Olsson, S. and Burns, P. (2000). Measuring distraction with a peripheral detection task. On-line paper. NHTSA Internet Distraction Forum. Available at www.nrd.nhtsa.dot..gov/departments/nrd-13/driver-distraction/welcome.htm