U.S. Department
of Transportation
National Highway
Traffic Safety
Administration

www.nhtsa.gov

DOT HS 812 077

November 2014

Detection Response Task (DRT) Evaluation for Driver Distraction Measurement Application

DISCLAIMER

Suggested APA Format Citation:

Ranney, T. A., Baldwin, G. H. S., Smith, L. A., Mazzae, E. N., & Pierce, R. S. (2014, November). *Detection response task evaluation for driver distraction measurement application.* (Report No. DOT HS 812 077). Washington, DC: National Highway Traffic Safety Administration.

Technical Report Documentation Page

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
DOT HS 812 077		
4. Title And Subtitle Detection Response Task (DRT) Evaluation for Driver Distraction Measurement Application		5. Report Date November 2014 6. Performing Organization Code NHTSA/NVS-312
7. Author(s) Thomas A. Ranney, G. H. Scott Baldwin, and Larry A. Smith, Transportation Research Center Inc.; Elizabeth N. Mazzae and Russell S. Pierce, National Highway Traffic Safety Administration		8. Performing Organization Report No.
9. Performing Organization Name and Address National Highway Traffic Safety Administration Vehicle Research and Test Center P.O. Box 37 East Liberty, OH 43319		10. Work Unit No. (TRAIS) 11. Contract or Grant No.
12. Sponsoring Agency Name and Address National Highway Traffic Safety Administration 1200 New Jersey Avenue SE. Washington, DC 20590		13. Type of Report and Period Covered Final Report 14. Sponsoring Agency Code

15. Supplementary Notes

The authors wish to thank Ed Parmer, Adam Andrella, Jodi Clark, Matt Hostetler, and John Martin for their contributions. Russell Pierce was principally involved in the identification of an appropriate measure of effect size, conducting effect size analyses/simulations, and provided editorial comments on the effect size section.

16. Abstract

Research was conducted to support development of NHTSA's Phase 3 Driver Distraction Guidelines for auditory-vocal driver-vehicle interfaces. **Experiment:** A single experiment was conducted in driving and non-driving test venues to evaluate the sensitivity of Detection Response Task (DRT) metrics to differences in attentional load. Three DRT variants were used: Head-mounted DRT (HDRT), Remote DRT (RDRT), and Tactile DRT (TDRT). A repeated-measures design required participants to complete all six combinations of DRT (3) and test venue (2). Secondary tasks included: 0-back, 1-back, and visual-manual radio tuning. Forty-eight participants provided two independent samples following the Phase 1 NHTSA Driver Distraction Guidelines selection criteria. Each 24-person sample had 6 participants (3 male, 3 female) in the following age ranges: 18-24, 25-39, 40-54, and 55+. Two identical stationary vehicles were used. One was connected to a fixed-base driving simulator where drivers maintained a constant following distance behind a lead vehicle. The other vehicle housed a non-driving venue in which participants performed DRT and secondary tasks without a concurrent driving task. Tasks were performed continuously for 3 minutes on each trial.

Results: Differences between test venues were more pronounced than differences between DRT variants. Response times were faster in the non-driving venue but differences between secondary task conditions were consistent across venues; radio tuning was associated with highest DRT performance degradation, followed by the 1-back and 0-back conditions, respectively. Non-driving hit rates were consistently higher than driving hit rates. A set of four planned comparisons and comparisons of effect sizes (ES) were used to evaluate metric sensitivity at different test durations. In the driving simulator, TDRT was slightly more sensitive than other DRT variants. In the non-driving venue, all three DRT variants provided comparable sensitivity for response time. Ceiling effects rendered hit rate data not sensitive to differences between conditions in the non-driving venue. A 2-minute data collection interval provided optimal sensitivity for testing. Based on (2-minute) small-sample ($N = 24$) testing in the driving simulator, both TDRT and HDRT detected all differences while RDRT performance was weaker. The TDRT had the highest level of test-retest reliability. For small sample non-driving venue testing, the 2-minute interval was also slightly better; however, there were no differences among DRT variants in sensitivity or test-retest reliability. Overall, the TDRT was consistently more sensitive than other DRT variants, albeit marginally, in the driving simulator. Using a non-driving venue requires reliance entirely on the response time metric. Differences between DRT variants were too small to identify a better performing DRT in the non-driving venue. Potential visual conflicts associated with RDRT and to a lesser extent with HDRT may create problems when used with visual-manual tasks in either venue.

17. Key Words	18. Distribution Statement
Driver performance, distraction, attention, detection response task (DRT), distraction guidelines, auditory-vocal driver-vehicle interfaces	Document is available to the public from the National Technical Information Service www.ntis.gov

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified	128	

Form DOT F 1700.7 (8-72) Reproduction of completed page authorized

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ACRONYMS / ABBREVIATIONS

AM	amplitude modulation; band used in radio tuning task
ANOVA	analysis of variance; statistical method for analyzing differences among means
Cohen’s d	standardized measure of effect size in which the average difference between each individual’s scores is divided by the standard deviation of the differences
CD	compact disc; media storage
CF	car following; driving task
DRT	detection response task; tool for assessing distraction
DS	driving simulator
ES	effect size
FM	frequency modulation; band used in radio tuning task
F Value	test statistic of an ANOVA
HDRT	head-mounted detection response task
HR	hit rate; DRT metric used to determine detection accuracy; same as hit rate
ICF	informed consent form; document describing test participant’s rights & responsibilities
ISO	International Organization for Standardization
IVIS	In-Vehicle Information System
LED	light emitting diode; used as target for HDRT and RDRT
Min / min	Minute
mph	Miles Per Hour
MULTTEST	SAS procedure to adjust probability values to correct for familywise error when multiple comparisons (family) are performed together
N	number; sample size
N-back	delayed response task in which participants listen to and repeat simple auditory stimuli; levels include 0-back and 1-back
NHST	null hypothesis significance testing; traditional statistical model using $p < .05$ as significance criterion
NHTSA	National Highway Traffic Safety Administration
OE	original equipment
OED	object and event detection
P, P Value	probability value associated with a statistical test representing the probability that the test outcome indicates a real world difference when one does not exist in reality (probability of Type I error)
PCorrect	proportion correct in DRT, same as hit rate
PDT	peripheral detection task, predecessor to the DRT
Pearson R	statistical measure of correlation between two variables
RDRT	remote detection response task
RT	response time, measure from DRT stimulus activation to participant button press
SAS	Statistical Analysis Software; software package used for performing statistical tests

SD	standard deviation; measure of variation in a set of data
SE	standard error; standard deviation of the sampling distribution of a statistic, representing the precision of the estimate of the statistic
STISIM Drive	driving simulator software for STISIM, a driving simulator created by System Technologies Incorporated
TDRT	tactile detection response task
TNO	Netherlands Organization for applied scientific research
t-test	statistical test to assess difference between two mean values
VRTC	Vehicle Research and Test Center; NHTSA’s test laboratory

EXECUTIVE SUMMARY

NHTSA continues to develop voluntary Guidelines to promote motor vehicle safety by discouraging the introduction and use of in-vehicle and portable electronic devices that are excessively distracting. The Guidelines are intended to ensure that driver interfaces are designed to minimize driver workload for tasks performed while driving. The Guidelines specify criteria and acceptance test protocols for assessing whether a task has a minimal impact on driving performance and therefore can be performed safely while driving. In 2013, NHTSA released the Phase 1 Driver Distraction Guidelines that address in-vehicle tasks performed using original equipment (OE) in-vehicle devices with visual-manual interfaces. NHTSA is currently working on Phase 2 of the Guidelines, which will address visual-manual tasks performed using portable and aftermarket devices. This report presents work that addresses issues related to Phase 3 of NHTSA's Guidelines, which will cover tasks performed using device interfaces capable of auditory-vocal interactions. Because the tasks covered under Phase 3 may pose different demands on drivers' attention than those covered in Phase 1, this research examined different metrics for assessing task conformance with the Phase 3 NHTSA Guidelines.

Ongoing work by the International Organization for Standardization (ISO) is evaluating Detection Response Tasks (DRTs) to determine how best to measure the degree to which drivers' attention is affected by secondary task demands. The results of the ISO evaluation will form the basis for an International Standard. The ISO initiative is consistent with the Phase 3 NHTSA Guidelines' acceptance test protocol development research objectives because the demands on driver attention represent a major component of the distraction potential of tasks performed with auditory-vocal interfaces. The NHTSA Guidelines development also involves establishing criteria that define acceptable performance. Therefore, the overall objectives of the present research were to replicate and extend the ISO research through the evaluation of DRTs and to provide information to help NHTSA establish a benchmark criterion of acceptable performance for tasks performed using device interfaces capable of auditory-vocal interactions.

Detection Response Tasks (DRTs) evolved from the Peripheral Detection Task (PDT), a visual signal detection task that has been widely used to assess the effects of task demands on driver attention. DRTs differ from the PDT primarily in the method of target presentation. They also incorporate controls to eliminate sources of unwanted variability in the two DRT metrics (response time and detection accuracy [hit rate]) due to changes in drivers' head positions and conflicts between the visual demands of secondary tasks and DRT task demands. The Head-mounted Detection Response Task (HDRT) consists of a single LED attached to a fixture worn on the head near the left eye, allowing the target to remain in the same position relative to the participant's eye position, independent of head position. The Tactile Detection Response Task (TDRT) uses an electrical vibrator (tactor) taped on the participant's shoulder. The Remote Detection Response Task (RDRT) is similar to the traditional PDT; a single LED is placed away from the participant in a single location near the central field of view. Thus, while the HDRT and TDRT both eliminate variability due to head position, the TDRT also eliminates all potential visual conflicts. In contrast, the RDRT does not incorporate either control.

Traditionally, DRTs have been used while participants are driving, either in a simulator or real driving environment. More recently, to simplify test requirements, DRTs are being evaluated in protocols in which the DRT and secondary tasks are performed without a primary (driving) component. Research results from studies using these approaches have not yet been published. Therefore, one objective of the present study was to evaluate the sensitivity of selected DRT variants in both driving and non-driving test venues. The second objective was to provide information to help NHTSA establish a benchmark criterion of acceptable performance for secondary tasks performed using auditory-vocal interactions.

These objectives were addressed in a single experiment in which 48 participants performed a set of secondary tasks with each of three DRT variants (HDRT, TDRT, RDRT) in two test venues (driving simulator, non-driving). Participants were recruited to provide two independent samples (N = 24) that met the NHTSA Phase 1 Guidelines recommendations. For each sample, six participants (three male, three female) were recruited in each of the following age ranges: 18 to 24, 25 to 39, 40 to 54, and 55 and older. The experiment was conducted using two stationary vehicles having the same interior features, with one connected to a fixed-base driving simulator and the other set up for a non-driving DRT venue.

To simulate the demands of tasks performed with auditory-vocal interfaces, the research used a delayed response task (n-back) in which participants listened to and repeated simple auditory stimuli (digits). The task allows mental workload to be systematically varied; 0-back represents a mild task demand and 1-back represents a moderate task demand. The n-back has no visual or manual components and thus allows a direct assessment of the effects of different levels of attentional demand on DRT metrics. In addition, to provide continuity with the research conducted to support the Phase 1 NHTSA Guidelines, visual-manual radio tuning was used as a secondary task in the present study.

Participants performed secondary tasks together with each DRT variant in both test venues. For all DRT variants, a single stimulus was presented repeatedly every 3-5 seconds. Participants responded as quickly as possible to each stimulus presentation via button press. The secondary task conditions included 0-back, 1-back, visual-manual radio tuning, and baseline (no secondary task). Thus, each participant completed four trials (one per secondary task condition) at each of the six DRT x Test Venue combinations. In the driving simulator venue, participants performed DRT and secondary tasks while performing a simple car-following task in a low-fidelity fixed-base simulator. In the non-driving test venue, participants performed DRT and secondary tasks with no concurrent driving task, while sitting in the driver's seat of an identical vehicle. The tasks were performed continuously during a 3-minute data collection interval on each trial.

Analyses based on four planned comparisons were performed to address three key questions:

1. Do differences exist among DRT variants (HDRT, TDRT and RDRT) that would make one preferable for use in testing?

The present results revealed only minor differences among the DRT variants. All three DRT variants were generally successful in detecting the differences defined at the outset of the study. Results based on the response time metric were consistently positive with all DRT variants in both test venues. TDRT revealed minor advantages in the driving simulator, including better sensitivity for detecting the most challenging planned comparisons and slightly better test-retest reliability. None of the DRT variants stood out in the non-driving venue.

Effect Size (ES) analyses revealed elevated RDRT response times in the radio tuning condition, reflecting uncontrolled variability due to conflicts between the RDRT and secondary task requirements when used with visual-manual tasks. HDRT and TDRT variants had smaller ES values for these differences, consistent with the effective operation of controls incorporated to minimize contributions from these variability sources.

2. Do differences exist between test venues (simulator, non-driving) that would make one preferable for testing?

The main difference between venues had to do with hit rate, which was consistently less sensitive to the targeted differences in the non-driving venue than in the driving simulator venue. Hit rates were higher and approached the ceiling defined by perfect performance in the non-driving venue to the point that this metric was not particularly useful in this venue. The response time metric in the non-driving venue was slightly more sensitive in detecting the targeted differences than in the driving simulator.

These results raise the question of whether the hit rate metric is essential to the analysis of DRT performance. If DRT testing requires a sensitive measure of hit rate, then the present results suggest that the non-driving test venue may be unsuitable for this purpose. In contrast, if hit rate is not essential, then the increased sensitivity and easier implementation offered by the non-driving venue would represent significant advantages over the driving simulator venue.

3. Do differences exist among data collection intervals of different durations?

Analyses performed on aggregated response time data found generally consistent patterns of results at intervals of 30 seconds, 1 minute, and 2 minutes; however larger ES values were associated with longer data collection intervals, with the biggest difference being observed between 2- and 1-minute intervals. The same pattern was observed for hit rate. These results suggest a progressive loss of sensitivity among the metrics at shorter data collection intervals.

Results based on planned comparisons with the small samples, provided support for using a 2-minute data collection interval. For simulator testing, the 2-minute interval facilitated better detection of differences targeted in the planned comparisons than the 1-minute interval among all DRT variants; however, this advantage was considerably smaller in the non-driving venue than in the driving simulator venue.

Additional analyses and discussion addressed a number of questions, including whether venues used for DRT testing require demonstrated validity in representing the demands of secondary tasks while driving, whether two valid performance metrics (response time and hit rate) are required, and the relation between DRT performance and safety. Questions in need of additional consideration that were not addressed in this work include determining the maximum acceptable level of attentional demand and the required metric sensitivity, which can be used to define the amount of attentional demand beyond a defined threshold level that is considered unacceptable.

Conclusions

The results support the following conclusions:

- Differences among DRT variants were small, but the TDRT was slightly more sensitive than the HDRT and RDRT when used in the driving simulator venue.
- The three DRT variants provided comparable sensitivity in the non-driving venue.
- The hit rate metric was generally less sensitive than response time in both test venues; consistently near-perfect target detection significantly reduced the sensitivity of this metric in the non-driving venue.
- The response time metric revealed slightly greater sensitivity in the non-driving venue than in the driving simulator venue.
- A 2-minute data collection interval provided optimal sensitivity for driving simulator venue testing, particularly for small-sample comparisons. A slightly shorter interval may be feasible in the non-driving venue without significant loss of metric sensitivity.
- The non-driving venue offers easier implementation plus slightly greater sensitivity for the response time metric. The hit rate metric provided insufficient sensitivity for consistently discriminating among secondary task conditions in this venue.
- The driving simulator venue provides a more valid representation of the concurrent demands of driving and secondary task performance. Both response time and hit rate metrics provided adequate sensitivity for discriminating among secondary task conditions in this venue.

1. INTRODUCTION

The National Highway Traffic Safety Administration (NHTSA) is developing voluntary Guidelines to promote safety by discouraging the introduction of excessively distracting devices in vehicles. The goal of these Guidelines is to ensure that driver interfaces are designed to minimize driver workload for tasks performed while driving. The Guidelines specify criteria and acceptance test protocols for assessing whether a task has a minimal impact on driving performance and therefore can be performed safely while driving. In 2013, NHTSA released the Phase 1 Driver Distraction Guidelines that address in-vehicle tasks performed using original equipment in-vehicle devices with visual-manual interfaces (NHTSA, 2013). The Phase 2 NHTSA Guidelines will address visual-manual tasks performed using portable and aftermarket devices. The present work addresses issues related to Phase 3 of NHTSA's Guidelines, which will cover tasks performed using device interfaces capable of auditory-vocal interactions. Because the tasks covered under Phase 3 may pose different demands on drivers' attention than those covered in Phase 1, this research examined different metrics for assessing task conformance with the Phase 3 NHTSA Guidelines.

The relation of driver distraction to inattention is embodied in the following definition: "Driver distraction is the diversion of attention from activities critical for safe driving to a competing activity," (US-EU Bilateral ITS Technical Task Force, 2010; see also Pettitt, Burnett, & Stevens, 2005). The following definition relates distraction to inattention: "Attention is most generally characterized as a combination of an activation level, which refers to how much attention is allocated to a given activity and the selectivity, which refers to how the attention is allocated among one or more activities." A technical working group of the International Organization for Standardization (ISO) is attempting to determine how best to measure the degree to which drivers' attention is affected by secondary task demands. As a starting point, they developed a conceptual framework for understanding driver attention (ISO, 2013). This model, which integrates ideas from several contemporary theories of attention, posits three levels of activity to which attentional resources can be allocated: (1) sensory/actuator resources; (2) perceptual/motor resources and (3) executive control. Sensory/actuator resources represent the lowest level structural resources, including the eyes, hands and feet. Perceptual/motor resources refer to the brain functions that control the perceptual/motor activities. Executive control, which may also be referred to as executive attention or supervisory control, refers to the higher-level operations such as decision making, sustaining information in working memory, mental computation, and interpretation of novel information. Allocation of executive control is generally a top-down process; it requires mental effort and is accessible to conscious awareness. With consistent practice, the executive control of tasks slowly becomes automatized, at least partially, and requires a lesser amount of mental effort.

Secondary tasks that require drivers to look at displays and physically manipulate controls have significant visual and manual components. Performing visual-manual tasks requires sensory/actuator resources (hands and eyes), perceptual/motor resources (controlling the hands and eyes), and some amount of executive control. Many contemporary secondary tasks can now be performed with interfaces that incorporate auditory-vocal (i.e., voice) interaction components. The design of auditory-vocal device/system interfaces is predicated on the assumption that if drivers can avoid looking away from the forward driving scene and avoid removing their hands from the steering wheel to interact with device controls, distraction effects will be minimized. This assumption is likely true, but applies primarily to the resources allocated to sensory/actuator and perceptual/motor task components. It does not consider the extent to which tasks may differ in their requirements at the third level of attentional activity, namely executive control or executive attention. In particular, some tasks may have comparable visual and manual demands but different levels of attentional demand. The development of DRTs is intended to address the need to assess the effects of increasing load on the executive aspects of attention.

Among distracted drivers, the diversion of attention from activities critical for safe driving will lead to situations in which the driver either misses or is slow in responding to events that occur on the roadway. Numerous studies have demonstrated that attentional distraction results in slowed response time (e.g., Horrey & Wickens, 2006; Caird, Willness, Steell, & Scialfa, 2008). Two primary methods have been used to assess this effect. One method, called object and event detection (OED), involves recording drivers' responses to unexpected hazards, such as a stopped vehicle in the roadway ahead (Lee, McGehee, Brown, & Reyes, 2002; Strayer, Drews, & Johnson, 2003). OED methods typically present realistic events to elicit the surprise inherent in drivers' responses to unexpected critical events. The second method is the set of tasks referred to as Detection Response Tasks (DRTs). DRT methods record drivers' responses to simple targets that are presented frequently. The use of surprise events in OED methods severely limits the number of presentations that can be made and thus the usefulness of this approach for testing protocols that require comparisons across multiple conditions. Although less realistic, DRT methods are supported by a long history of research demonstrating their sensitivity to differences in levels of attentional demand (Victor, Engström, & Harbluk, 2009). The strong empirical foundation together with the practical advantages of DRT methods led to the ISO decision to select the DRT as the most promising method for assessing differences in the potential for distraction due to differences in attentional demand between tasks (ISO, 2013). The ISO initiative will lead to an International Standard (ISO, 2014).

1.1 Detection Response Tasks (DRTs)

DRT evolved from what had been referred to as the Peripheral Detection Task (PDT). The original PDT was based on the work of Miura (1986), who studied the effects of driving task demand on target detection using lights projected onto a vehicle windshield. Drivers were typically asked to respond as quickly as possible via button press when they detect a target. Response time and detection accuracy were the primary metrics. Miura found that with increasing traffic demands response times to stimuli increased and detection accuracy to targets projected at greater eccentricities from the forward view decreased. Early studies, reviewed by Victor, Engström, and Harbluk (2009), found the PDT to be sensitive to workload but not consistently sensitive to differences in target location. The failure to find evidence of visual tunneling with increasing workload led to some confusion about the theoretical basis of the test, but its sensitivity to the effects of visual-attentional and purely attentional secondary task demands fostered continued use and development.

The PDT was used in a number of NHTSA research studies involving both instrumented vehicle and simulator implementations. Initial studies (Ranney, Mazzae, Baldwin, & Salaani, 2007; Ranney, Harbluk, Smith, Huener, Parmer, & Barickman, 2003) used a dashboard-mounted version following specifications presented by Harms and Patten (2003), which required drivers to detect and respond to simple targets reflected on the windshield (see Figure 1). More recently, several variants of the PDT have emerged. Variants differ primarily in the method of target presentation and the number of targets; the frequency of targets and the method of response (micro-switch attached to index finger) have generally remained consistent. An overview of selected variants is presented in the following sections.

Figure 1. Original PDT Location With Single LED Activated

1.1.1 Visual Target Detection Task

The visual target detection task was a modification of the original PDT designed to take advantage of the flexibility in target presentation offered by a driving simulator display (Ranney, Baldwin, Parmer, Martin, & Mazzae, 2011; Ranney, Baldwin, Parmer, Domeyer, Martin, & Mazzae, 2011). Instead of light emitting diodes (LEDs) reflected off the windshield, the targets were computer-generated, red-colored circles intended to approximate the size of the reflected LEDs in the traditional PDT. The presentation of targets on the roadway display eliminated the constraints associated with reflected LEDs and allowed a wider range of target locations than the original PDT. For example, the Visual Target Detection Task, shown in Figure 2, presented targets at six different locations in the roadway display.

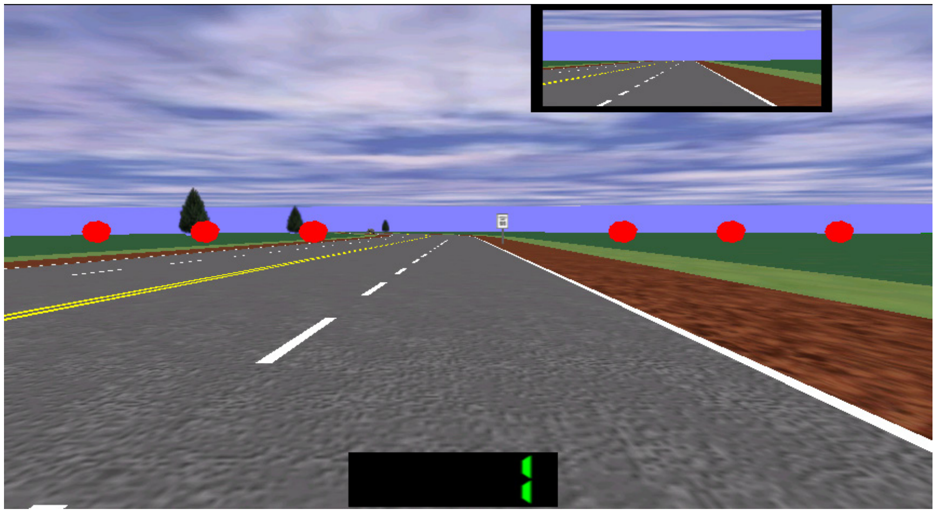


Figure 2. Modified Visual Target Detection Task

The earlier PDT versions presented targets at fixed external (remote) locations either as reflections on the vehicle windshield or on a roadway display. The use of fixed external target locations was thought to provide a degree of face validity in the context of driving, as it can be imagined that visual targets in the roadway display are generic representations of hazards that drivers would routinely be required to interpret and respond to while driving. However, recent theoretical work has better defined the underlying behavioral mechanisms that contribute to DRT performance (Engström, 2010). According to this analysis, DRT variants attain much of their sensitivity because, unlike in OED tasks, the targets are predictable, expected by drivers, and require no interpretation. Response metrics associated with task versions that used fixed external target locations included unwanted variability due to differences in head position at the time targets were presented. Newer DRT variants introduced design features to fix the target location relative to the driver's head position to eliminate this unwanted variability.

1.1.2 Head-mounted Detection Response Task

The Head-mounted Detection Response Task (HDRT) was developed by TNO Research Institute in the Netherlands (van der Horst & Martens, 2010). It consists of a single LED attached to a fixture worn on the head and positioned to be to the left and slightly above the participant's left eye (see Figure 3). Mounting the LED on the head allows the target to remain in the same position relative to the driver's eye position, independent of head position, which eliminates uncontrolled response variability due to the differences between head position and target location when targets are positioned in the driving environment.

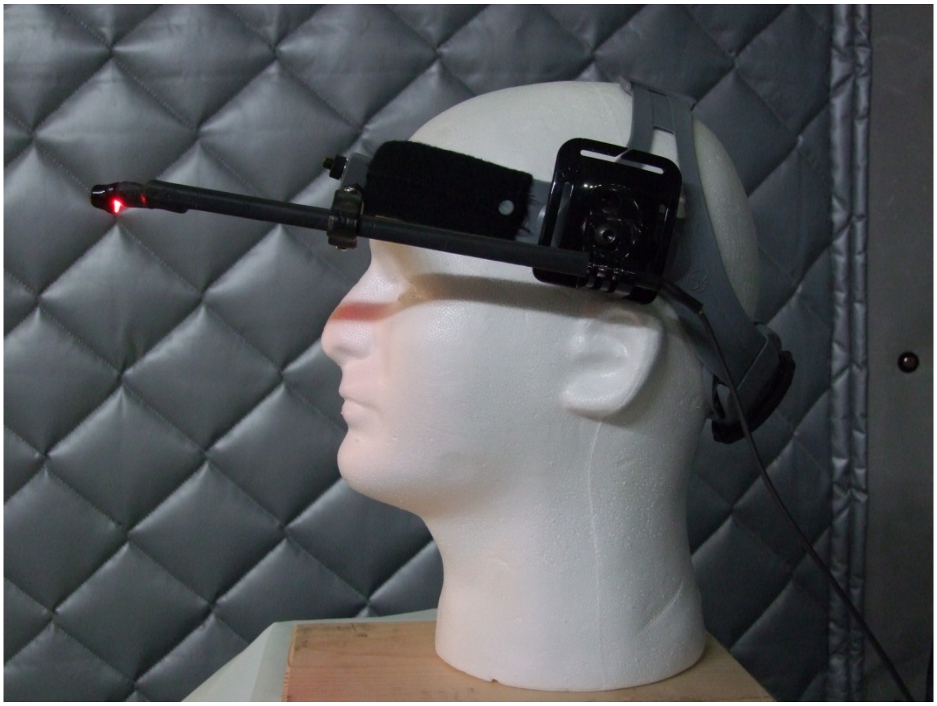


Figure 3. Head-mounted DRT

1.1.3 Tactile Detection Response Task

The tactile detection response task (TDRT) consists of a small electrical vibrator (tactor) taped on the left shoulder of the participant, as shown in Figure 4. Like the HDRT, the target location remains fixed relative to the driver, thus eliminating response variability due to head and eye gaze positions. In addition, by eliminating the potential conflict between the detection of visual targets and the visual demands of driving, the TDRT variant provides an additional level of control that helps isolate the effects of increasing attentional load. One would therefore expect the TDRT to be the purest among the DRT variants as a measure of attentional demand (Engström, 2010). One potential limitation of this approach is that the use of a vibration stimulus is much less well-studied in the context of driving than the use of visual stimuli. Although some research has explored the effects of different locations, the placement of the tactor on the participant's body may have some effect on response speed.

Figure 4. Tactile DRT

1.1.4 Remote Detection Response Task

The remote detection response task (RDRT) consists of a single LED placed remotely in a fixed location near the driver's central field of view (see Figure 5). This DRT variant differs from the HDRT and TDRT in that the target location is not fixed relative to the driver. Rather, this variant uses the traditional approach of fixing the target location external to the driver. It thus reintroduces the possibility that target detection speed and accuracy will be affected by targets not being visible immediately upon activation when the driver is looking away from the forward roadway view. Like the newer DRT variants, it uses a single target, following the above-mentioned conclusion that the PDT was not consistently sensitive to differences in target eccentricity (Victor, Engström, & Harbluk, 2009). This variant represents an intermediate step between the traditional PDT/DRT versions that used multiple external targets and the newer variants (HDRT and TDRT), which use single targets presented at fixed locations relative to the driver. The failure of the RDRT to control for head position differences at target activation is expected to reduce metric sensitivity for detecting differences between tasks relative to the HDRT or TDRT, but this difference will only be apparent for visual-manual tasks, which require drivers to look away from the roadway scene.

Figure 5. Remote DRT

1.1.5 DRT Without a Concurrent Driving Task

Traditionally, DRTs and their predecessors have been presented while participants are driving, either in a simulator or a real driving environment. More recently, in an attempt to simplify test requirements, DRTs are beginning to be evaluated in protocols in which the DRT and secondary task are performed without a primary (driving) component. Alternatively, the DRT has been paired with tasks intended to represent simple surrogates of driving task components. Examples include passive, semi-static video presentations of moving roadway images without interactive controls and displays, or tracking tasks with a simple two-dimensional display intended to represent an abstraction of some aspect of vehicle control. Research results with these approaches have not yet been published; however the preliminary results from ISO testing indicate that the approaches that do not include concurrent driving are thought to provide results that are generally comparable to those that use the traditional dual-task protocol involving a driving task. The major benefit associated with non-driving DRT variants is that testing could be accomplished with simpler setups that do not require an interactive driving simulator.

1.2 Research Objectives

Recent theoretical advances (e.g., Engström, 2010) have clarified the behavioral mechanisms underlying DRT performance, providing a basis for differentiating DRT target detection from OED methods. Based on this work, there is emerging consensus that DRTs assess the attentional demands of secondary tasks. This clarification has been coincident with efforts to redesign the DRT, which have improved metric sensitivity by eliminating unwanted variability that was inherent in earlier versions.

DRT variants have been examined in a coordinated set of experiments conducted in support of ISO standard development (ISO, 2013). The ongoing ISO initiative is consistent with the Phase

3 NHTSA Guidelines' acceptance test protocol development research objectives because the demands on driver attention represent a major component of the distraction potential of tasks performed with auditory-vocal interfaces. Therefore, one objective of the present research was to assess the feasibility of using DRT metrics as part of Phase 3 NHTSA Guidelines testing. Specifically, this work was conducted to replicate and extend the ISO research through the evaluation of different DRT variants, including the HDRT, TDRT and RDRT. Following the ISO work, the second study objective was to evaluate the selected DRT variants in both driving and non-driving test venues.

Existing ISO research and DRT research more generally have not attempted to establish a threshold to define an acceptable level of attentional demand. However, because NHTSA guidelines would likely include such a threshold, one objective of the present work was to provide information that will help NHTSA establish a benchmark criterion of acceptable performance for tasks performed with auditory-vocal interfaces. NHTSA's Phase 1 Guidelines used visual-manual radio tuning as a benchmark to define the threshold of acceptable visual-manual distraction. In the present experiment, the attentional demands of visual-manual radio tuning were compared to those of standardized auditory-vocal calibration tasks to provide a preliminary assessment of the feasibility of using this level of demand as a threshold for acceptable attentional demands. Establishing a benchmark level of acceptable attentional demand is a complex task that will require additional work. Accordingly, the present experiment was expected to provide no more than a preliminary indication of the level of acceptable attentional demand.

2. METHOD

2.1 Approach

The study objectives were addressed in a single experiment. To facilitate the comparison between driving simulator and non-driving test venues, the experiment was conducted using two stationary vehicles having the same interior features. One vehicle was connected to a fixed-base driving simulator that engaged drivers in a car-following scenario while performing secondary tasks and DRT. The second vehicle housed a non-driving venue; participants performed DRT and secondary tasks without a concurrent driving task. Three DRT variants (HDRT, TDRT, and RDRT) were tested in each venue. The driving simulator scenario used in this research was designed to be consistent with the specifications presented in NHTSA's Phase 1 Guidelines.

Because the focus of the experiment was on assessing the DRT metrics' sensitivity for detecting differences in the attentional demands of secondary tasks, the experiment included tasks designed to isolate the effects of attentional demand. The n-back task, a verbal delayed digit recall task (see Section 2.6.2) selected by ISO for this purpose, was used in this experiment. The n-back task has multiple levels of task difficulty, which allow the attentional demand to be systematically varied. It also has the advantage of being externally paced, which ensures a consistent level of task demand over time and thus eliminates unwanted variability due to individual differences in task completion rate.

Visual-manual radio tuning was included in the experiment to provide continuity with the Phase 1 NHTSA Guidelines test protocol recommendations and to allow assessment of whether the different DRT variants provide comparable information when used to assess secondary tasks with different interfaces.

2.2 Experimental Design

The experiment consisted of a repeated-measures within-subjects design in which all participants completed all treatment combinations. Independent variables included DRT variant (HDRT, TDRT, and RDRT), test venue (driving simulator, non-driving) and secondary task (0-back, 1-back, and visual-manual radio tuning).

In addition to the 18 combinations of DRT/test venues and secondary tasks, each DRT/test venue had a baseline condition with no secondary tasks. The simulator baseline involved car following plus DRT; the non-driving baseline involved DRT performance alone. Thus, each participant completed 24 data collection trials, as summarized in Table 1.

Table 1. Secondary Task by DRT by Test Venue Combinations

Trial	Secondary Task	DRT	Test Venue
1	0-back	Head-mounted (HDRT)	Static (Non-Driving)
2	1-back		
3	Radio tuning		
4	None (baseline)		
5	0-back	Tactile (TDRT)	
6	1-back		
7	Radio tuning		
8	None (baseline)		
9	0-back	Remote (RDRT)	
10	1-back		
11	Radio tuning		
12	None (baseline)		
13	0-back	Head-mounted (HDRT)	Driving Simulator
14	1-back		
15	Radio tuning		
16	None (baseline)		
17	0-back	Tactile (TDRT)	
18	1-back		
19	Radio tuning		
20	None (baseline)		
21	0-back	Remote (RDRT)	
22	1-back		
23	Radio tuning		
24	None (baseline)		

The order of presentation for all three factors (test venue, DRT variant, and secondary task condition) was counterbalanced so that an equal number of participants had each factor in each possible position. In particular, half of the participants completed the static venue trials first; the other half completed the simulator trials first. Within each venue, an equal number of participants had each order of DRT variants. Within each DRT variant, the order of secondary task conditions was balanced using two digram-balanced Latin squares, in which each position 1-4 and each pairing 1-2,2-3,3-4, etc. was represented an equal number of times. The effect of the counterbalancing was to spread order effects equally across the entire matrix.

The data collection interval for each trial was 3 minutes. For all DRT variants, a single stimulus was presented at temporal intervals selected randomly from a uniform distribution of times between 3 and 5 seconds (onset to onset). The stimulus remained active for a maximum of 1 second; if the participant detected the target and responded during this interval, the stimulus was extinguished immediately after the response (ISO, 2013).

2.3 Participants

Forty-eight drivers participated in this experiment. They were recruited according to the Phase 1 NHTSA Driver Distraction Guidelines specifications. For each group of 24 participants, 6 participants were recruited in each of the following age ranges: 18 to 24, 25 to 39, 40 to 54 and 55 plus. An equal number of men and women were included in each age range. Participants were in good general health, had a valid driver's license with no self-reported vision or hearing problems, drove at least 3,000 miles in the last year, claimed to have experience using a wireless phone while driving, and were unfamiliar with the technology being evaluated.

Data for this experiment were collected during July and August 2013.

Two participants reported significant motion sickness during the experiment. Of these, one was able to continue and complete the experiment following a rest, while the second was unable to complete the protocol and was replaced.

2.3.1 Recruitment

Participants were recruited using Web-based networks such as Craigslist and through advertisements placed in a local newspaper in Ohio, the Marysville Journal-Tribune. To facilitate recruitment, an online application allowed participants to complete the screening questionnaire online. Recruitment materials can be found in Appendix A.

2.3.2 Payment for Participation

Participants were compensated for their participation according to criteria developed by NHTSA. Compensation for participation consisted of the total of two amounts: (1) Base pay for participation ($42/hour), and (2) mileage reimbursement for travel to and from the test facility ($0.545 per mile).

2.4 Apparatus

For the non-driving test venue, the setup consisted of the following main components: a production vehicle (2010 Toyota Prius V), an Intel Pentium 4 laptop computer to control the DRTs and secondary task stimulus information, and a data acquisition system to collect DRT metrics and video from multiple camera locations.

For the driving simulator test venue, the experiment used the fixed-base driving simulator located at NHTSA's Vehicle Research and Test Center (VRTC). Components of the fixed-base simulator included the Prius production vehicle, an Intel Pentium 4 computer, a ceiling-mounted Infocus model LP815 digital projector (1024 x 768) positioned above the vehicle, and a forward projection screen (10 feet x 10 feet). The screen was located approximately 186 inches forward of an average driver's eye point. The STISIM Drive simulator software Version 2.06.03 was used. A separate computer was used to generate stimulus information for each secondary task. Additional details on the laboratory setup are presented in Ranney, Baldwin, Parmer, Martin, and Mazzae (2011).

In the simulator, a vehicle data acquisition system was configured to collect steering wheel, brake and throttle position inputs. The data acquisition system also provided time synchronization for all data, which included DRT data, video from multiple camera locations, and STISIM driving data.

The primary data collected by the data acquisition systems in both venues are shown in Table 2.

Table 2. Primary Data Collected in Both Venues, DRT Metrics

Data	Description	Units
Target Activation Time	Time relative to beginning of trial that DRT target was activated	Seconds
Target Response Time	Time relative to the beginning of the trial that DRT response was made	Seconds

The DRT performance metrics computed from these times are described in section 2.8.

The HDRT apparatus consisted of a single LED mounted on a 21 cm arm that was attached to a suspension hard-hat liner. The LED was positioned to be approximately 20 degrees to the left and 10 degrees above a participant's left eye (see Figure 3). Following the ISO draft standard (ISO, 2013), the LED was red with a 5 mm diameter. It had a dominant wavelength of 626 nm and a luminous intensity of approximately 0.055 cd.

The TDRT apparatus consisted of a small electrical vibrator (tactor) taped on the participant's left shoulder near the clavicle as shown in Figure 4. Following the ISO draft standard (ISO, 2013), the TDRT had a diameter of 10 mm, weight of 1.2 grams, speed of 12,000 rpm, and vibration amplitude of 0.8 G.

The RDRT apparatus consisted of a single LED placed remotely in a fixed location near the driver's central field of view (see Figure 5). Following the ISO draft standard (ISO, 2013), the RDRT used a red LED with a dominant wavelength of 626 nm, and was 5 mm in diameter. The ISO brightness specification was too bright for our test venues, resulting in a glow that could be detected uniformly independent of head position. To eliminate this glow, the brightness was reduced slightly.

All DRT variants used a micro-switch attached to the participant's left index finger, as shown in Figure 6.

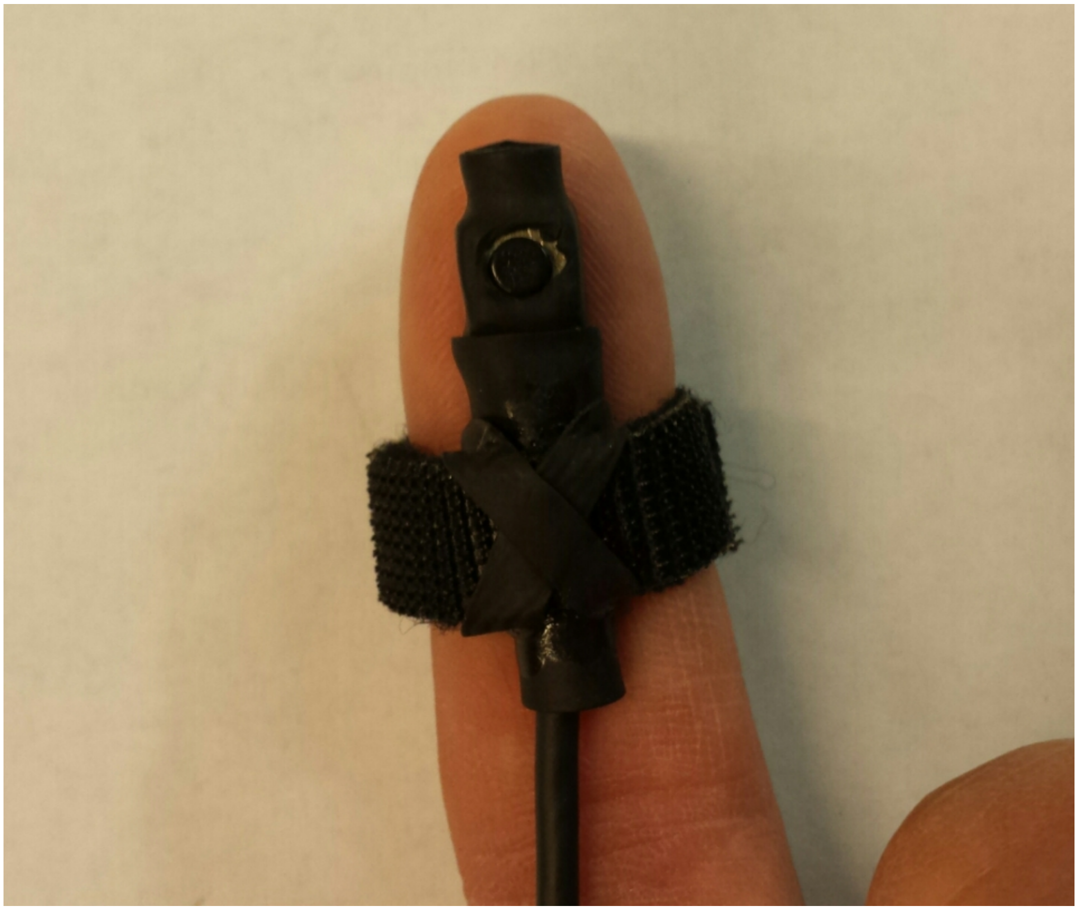


Figure 6. DRT Response Button

2.5 Driving Task

The driving task consisted of a simple car-following task on a straight 4-lane undivided roadway with no intersections. The lead vehicle maintained a constant speed of 50 mph. Following the Phase 1 NHTSA Guidelines test protocol, this task required participants to maintain a constant following distance of approximately 70 meters (220 feet). Feedback regarding maintenance of following distance was provided to the test participants after each trial. Other than the lead vehicle, there was no traffic present.

2.6 Secondary Tasks

The following secondary tasks were used in the experiment:

- Radio tuning (benchmark)
- 0-back
- 1-back

Details of each secondary task as implemented here are provided in the following sections.

2.6.1 Radio Tuning

Radio tuning is generally considered acceptable for performance while driving and was the benchmark level of acceptable demand for the visual metrics defined in the Phase 1 NHTSA Guidelines. Visual-manual radio tuning was included in this study to facilitate comparisons of DRT performance degradation with auditory-vocal (voice-based) tasks. Radio tuning tasks were performed using the original equipment (OE) in-vehicle information system in the test vehicles.

The radio tuning task had the following steps: (1) select the "audio" function of the built-in stereo; (2) select the frequency band by pressing the AM or FM button; and (3) use the tuning knob to adjust the frequency. When performed repeatedly, as required by the DRT protocol, an additional step was required following each task instance to return the information system to the predefined starting point. Radio tuning task training instructions are presented in Appendix B.

2.6.2 N-back (0-back & 1-back)

N-back is a verbal response delayed digit recall task in which a participant listens to and repeats a sequence of recorded single digits according to one of several specific rules (e.g., 0-back or 1-back) (Mehler, Reimer, & Dusek, 2013). The participant is required to respond after each presentation with either the same digit that was just presented (0-back, easy condition) or with the one that was previously presented (1-back, more difficult condition). Digits are presented at a predetermined rate (external pacing) to create a consistent load over the entire data collection interval. Example sequences for the 0-back and 1-back conditions are shown in Table 3.

Table 3. N-back Stimulus and Response Sequence

Task	**Digit presented**	**3**	**2**	**6**	**7**	**1**
0-back	Correct response	3	2	6	7	1
1-back	Correct response	-	3	2	6	7

In this table, the task sequence moves from left to right over time. The experimenter's presentation, which is the same for both conditions in this example, is shown in the top row. The correct response to be said aloud by the participant is presented in the rows beneath the digit presented. The 1-back task differs from the 0-back task in that it places a greater burden on working memory, which according to the conceptual model presented above represents one attribute of executive control. Complete n-back task training instructions can be found in Appendix B.

2.7 Procedure

Each participant completed a single session, which lasted approximately 7 hours. Upon arrival, the participant was asked to read and sign the Participant Informed Consent Form (Appendix C), thereby giving informed consent to participate in the study. The participant was then given an overview of the test venues (see Appendix B), before proceeding to the lab. In the lab, the participant was assigned to one of the two venues and was given an overview of the test vehicle including vehicle controls and displays. Once the participant was comfortable in the seat, training began, first with a description of how task performance feedback would be provided throughout the experiment.

The experimenter described each secondary task and provided detailed task instructions (see Appendix B). The experimenter then demonstrated each task after which practice was encouraged until the participant felt comfortable performing the task. Typically, two practice trials were needed for each secondary task. Participants were encouraged to ask for information to be repeated if they forgot some instruction. They were also instructed that data entry errors should be corrected before moving on. Participants were then introduce to and trained on the DRT variants.

In the simulator venue, participants were given driving task instructions and familiarization drives to practice the car-following task. (Additional details of the procedures and instructions are presented in Appendix B.)

When training was complete and the participant was comfortable with each of the tasks, the participant completed the 12 data collection trials in the first venue in the order specified by the experimental design. Each data collection trial was preceded by a practice trial with the same combination of tasks. Each task trial lasted approximately 3.5 minutes, which included 3 minutes of continuous data collection. Pre-recorded auditory stimuli were used for the radio tuning and n-back trials. When the data collection trials in the first assigned test venue were complete, the participant had the opportunity to take a lunch break. In the second test venue, the participants completed the remaining 12 data collection trials. Training for secondary tasks and DRT was abbreviated in the second venue since these tasks were identical to those used in the first venue. The data collection procedure for each trial included a practice trial followed by the data collection trial.

Following completion of the simulator data collection trials, the participant was assessed for simulator-induced motion sickness (see Appendix D) to determine if rest was required before moving on to the next venue or driving home. At the completion of both venues (all 24 data collection trials), the participant was given compensation and an opportunity to ask questions. After all questions were answered, the participant was accompanied by the experimenter to his or her personal vehicle and departed the research site.

The experimenters were positioned at a control station behind the vehicles during data collection. Communication with the participant was accomplished by a speaker and microphone system in the simulator venue, and directly by walking up to the driver's window in the non-driving DRT venue.

2.8 DRT Performance Metrics

DRT performance was assessed using two primary metrics, proportion of targets correctly detected (hit rate) and mean response time (RT). Hit rate was defined as the number of targets correctly detected (hits) divided by the total number of targets presented in a data collection interval. The mean response time was computed as the mean of all correctly detected targets during the same interval. Hits were defined as responses recorded between 100 and 2500 milliseconds following stimulus onset (ISO, 2013). This range is intended to allow inclusion of responses made when the participant does not detect the target until the end of the presentation interval and to eliminate spurious responses and/or attempts by participants to "game" the task by responding randomly. Accordingly, responses shorter than 100 milliseconds or longer than 2500 milliseconds were excluded from analysis.

Data obtained during the 3-minute data collection intervals were segmented in various ways to assess the effect of the number of target presentations on metric sensitivity. Differences between conditions were tested using data collection intervals of 15 and 30 seconds, as well as 1, 2, and 3 minutes.

2.9 Overview of Data Analysis

Analyses were performed to identify the most sensitive and reliable DRT variants in each test venue. Initial analyses were conducted using data from all (N = 48) participants. The large-sample analyses were intended to provide a "ground truth" against which the small-sample (N = 24) analyses could be compared. The small samples were the two independent 24-person samples created according to Phase 1 NHTSA Guidelines specifications, which together comprise the large sample. Initial analyses examined the stability of the two metrics over successive segments of 15 and 30 seconds. Omnibus analyses were conducted using data aggregated across all conditions. Separate analyses were performed for different data collection intervals because the intervals were overlapping.

A set of four planned comparisons was defined as the basis for comparing the sensitivity of the different DRT variants in each test venue. These included two diagnostic comparisons and two benchmarking comparisons. The rationale for the selected comparisons is presented in Section 3.6. The planned comparisons were initially done with the full sample using different data collection intervals and then with the small samples to examine the sensitivity in situations similar to those used in NHTSA Guidelines task acceptance testing.

Following these statistical comparisons, a set of analyses was presented that considered the effect sizes (ES) associated with the four planned comparisons at different data collection intervals for each combination of test venue and DRT variant. Additional analyses were performed to address a set of ancillary questions, including test-retest reliability based on the outcomes associated with the two independent 24-person samples and a comparison of the effects of using parametric versus non-parametric statistical approaches.

3. RESULTS

A 3-minute data collection interval was used for each task trial in this experiment. This duration is consistent with previous work and reflects the assumption that an extended interval is necessary to provide a sufficient number of target presentations to ensure stable summary measures, which include response time and hit rate. There were approximately 45 target presentations during each 3-minute data collection interval. Recent ISO work, however, has focused on shorter intervals, including intervals of 1 to 2 minutes, with some parties advocating for even shorter intervals. One objective of the present analysis therefore was to determine the shortest interval necessary to ensure stable performance measures. The analyses presented herein were conducted using both the full 3-minute interval and subsets of shorter durations. Unless otherwise stated, all such intervals are referred to as data collection intervals. All data collection intervals use the same starting point, which means that the data collection intervals of different durations are overlapping and therefore not independent. Figure 7 presents mean response time over the 3-minute data collection interval for each task performed under each combination of DRT and test venue. Standard error (SE) values, shown as error bars in the figure, represent the precision of the estimated mean value. As described in Section 2.2, the baseline conditions differ in the two test venues.

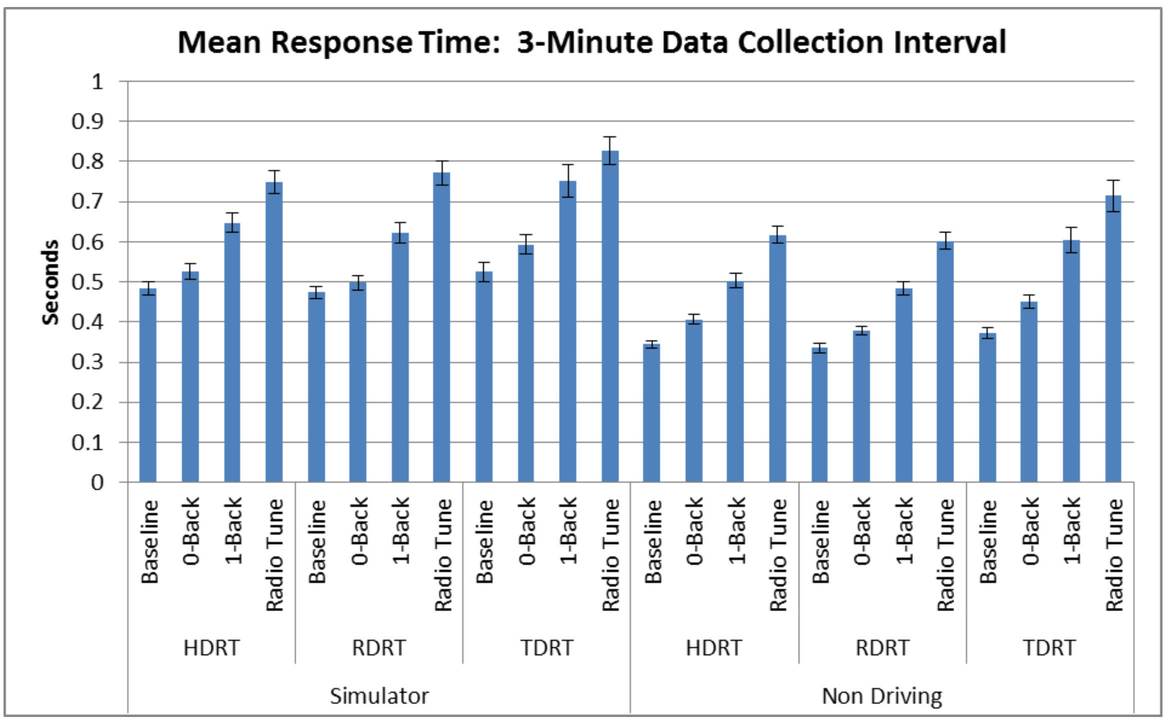


Figure 7. Mean Response Time (± SE) by Test Venue, DRT, and Secondary Task: 3-Minute Data Collection Interval

The order of secondary task conditions was consistent across DRT variants; minor variations are apparent in the magnitude of differences among the conditions across DRT variants. The baseline condition (no secondary task) was consistently associated with the fastest response times, as would be expected. The 0-back, 1-back and radio tuning conditions were associated with slower response times, respectively, indicating successively increasing DRT performance impairment. Response times were consistently faster in the non-driving test venue, as would be expected due to the absence of a concurrent driving task.

Figure 8 presents the mean values for hit rate for each combination of test venue, DRT, and secondary task, using the 3-minute data collection interval.

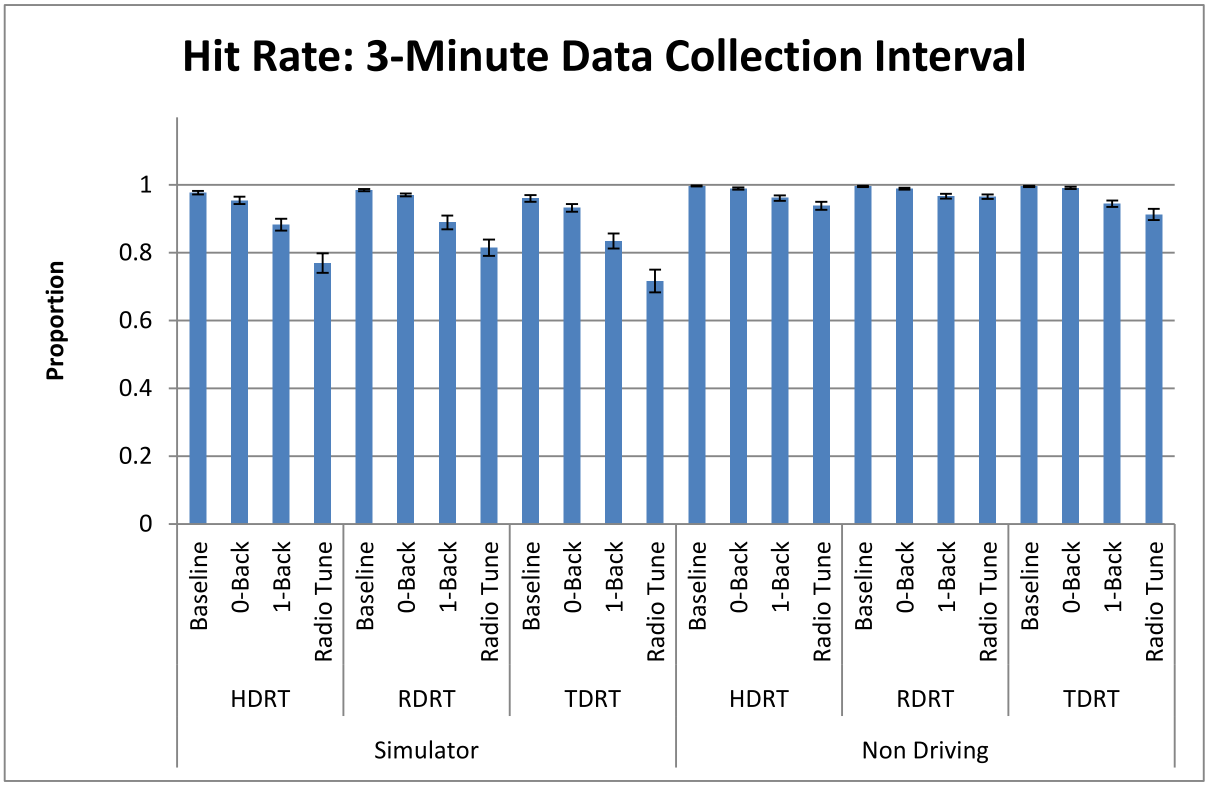


Figure 8. Mean Hit Rate by Test Venue, DRT, and Secondary Task: 3-Minute Data Collection Interval

Based on this figure, it is apparent that the pattern of mean hit rate values across secondary task conditions is generally different in the two test venues. The mean hit rate values in the simulator conditions revealed a pattern that is consistent with the mean response time pattern in the ordering of secondary task conditions; DRT performance was best in the baseline condition and worst in the radio tuning condition. However, this pattern, while apparent, is significantly attenuated in the non-driving conditions. Without a concurrent driving task, drivers were better able to detect the DRT targets in all task conditions. As a result, hit rate exhibited considerably less sensitivity to differences between conditions. These effects were examined in greater detail; statistical test results are presented in the following sections.

Figure 9 presents the mean response time values for each combination of secondary task, DRT and test venue for data collection intervals of 1, 2, and 3 minutes. The corresponding hit rates are presented in Figure 10.

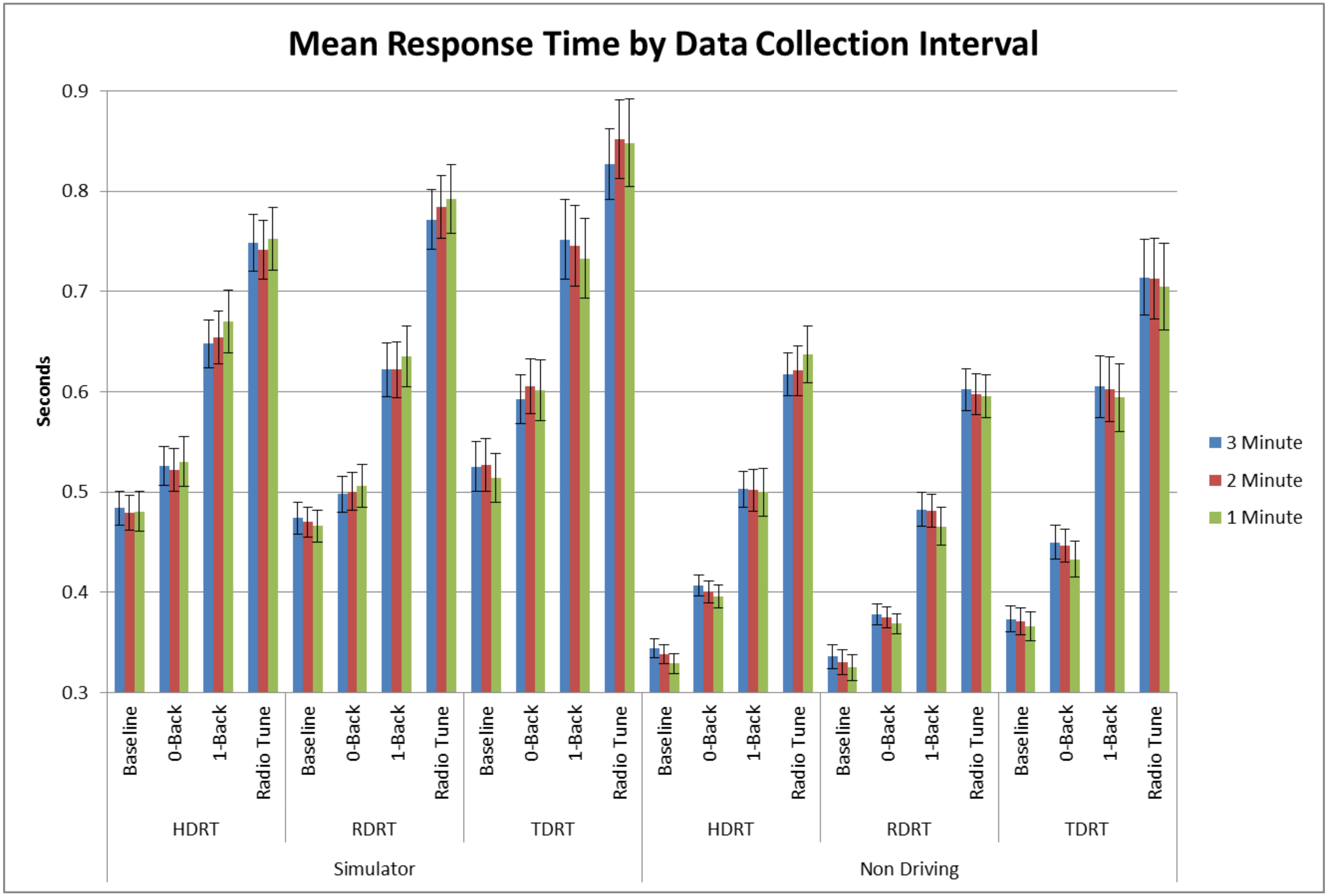


Figure 9. DRT Response Times (±SE) by Secondary Task, DRT, Test Venue, and Data Collection Interval

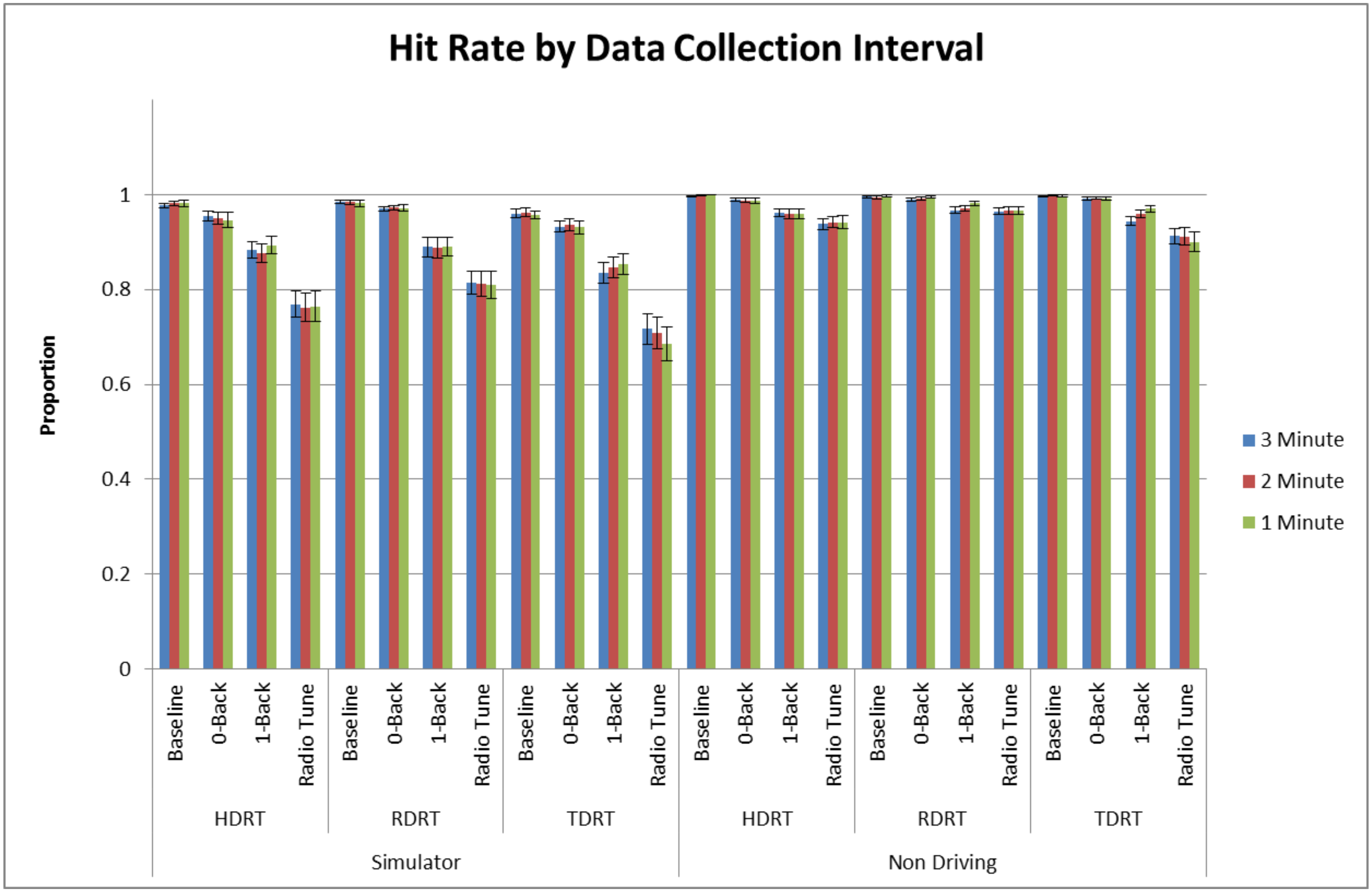


Figure 10. DRT Hit Rate by Secondary Task, DRT, Test Venue, and Data Collection Interval

As shown in Figure 9, the basic trends for mean response time values remain consistent across the three data collection intervals. Differences between the driving simulator and non-driving venues are consistent across the DRT conditions and data collection intervals, suggesting a fundamental difference in the detection task between the two test venues. Specifically, as indicated above, faster response times associated with non-driving task conditions indicate that targets were generally much easier to identify when there was no concurrent driving task.

Among the DRT conditions in the non-driving test venue, the response times associated with the TDRT conditions were slightly slower than those in the other two DRT conditions. TDRT response times in the radio tuning task condition were slower than for the other two DRT conditions. RDRT response times were slightly faster than the HDRT response times; target detection was generally accomplished more quickly in the RDRT condition than in the other two DRT conditions.

In the simulator venue, TDRT response times were also generally slower than those for the other DRT variants. RDRT response times associated with 0-back and 1-back conditions were generally faster than comparable response times in the other two DRT conditions. TDRT response time means exhibited the largest absolute differences between baseline, 0-back and 1-back conditions relative to the other two DRT conditions.

The pattern of hit rates shown in Figure 10 also appears generally consistent across data collection intervals of different duration. Some task conditions reveal slight increases in variability at the shorter data collection intervals, which may suggest reduced statistical power with shorter data collection intervals. Statistical tests were performed and are reported in subsequent sections.

3.1 Response Time by 30-Second Segments

As background for the examination of patterns at different time intervals, the 3-minute data collection intervals were separated into six 30-second data segments. At approximately 15 target presentations per minute and assuming detection accuracy of 80 percent for driving conditions and over 90 percent for the non-driving venue, each 30-second segment summarizes data from approximately 6 targets for driving trials and 7 targets for non-driving trials. Figure 11 through Figure 16 present the mean response times for each combination of DRT and test venue.

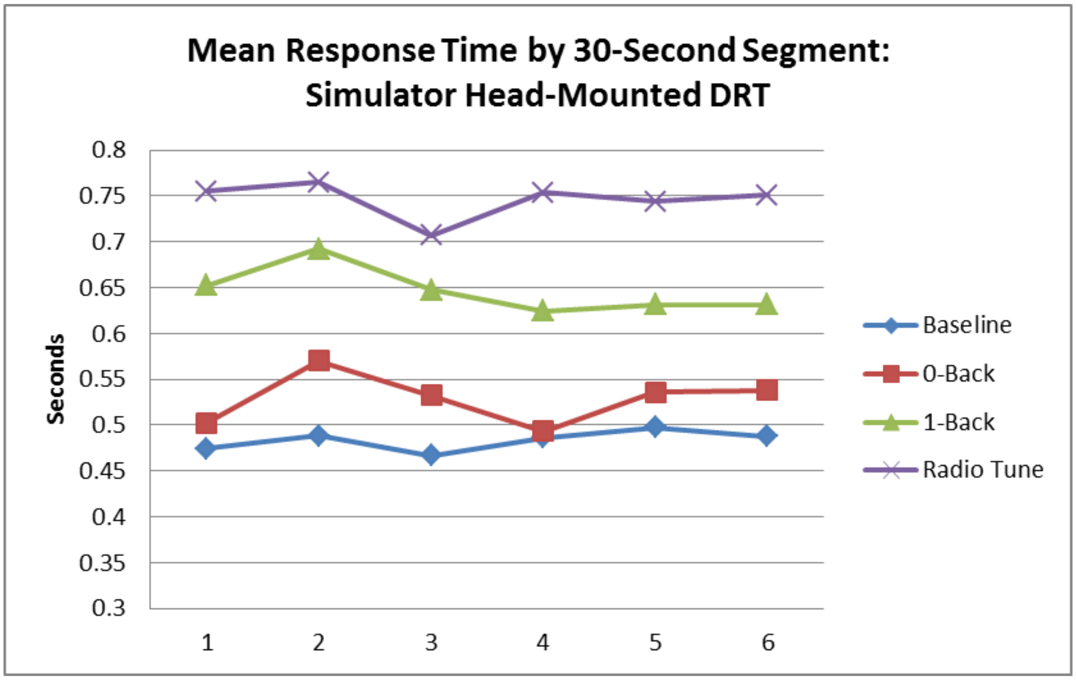


Figure 11. Mean Response Time by 30-Second Segment: Head-mounted DRT in Driving Simulator

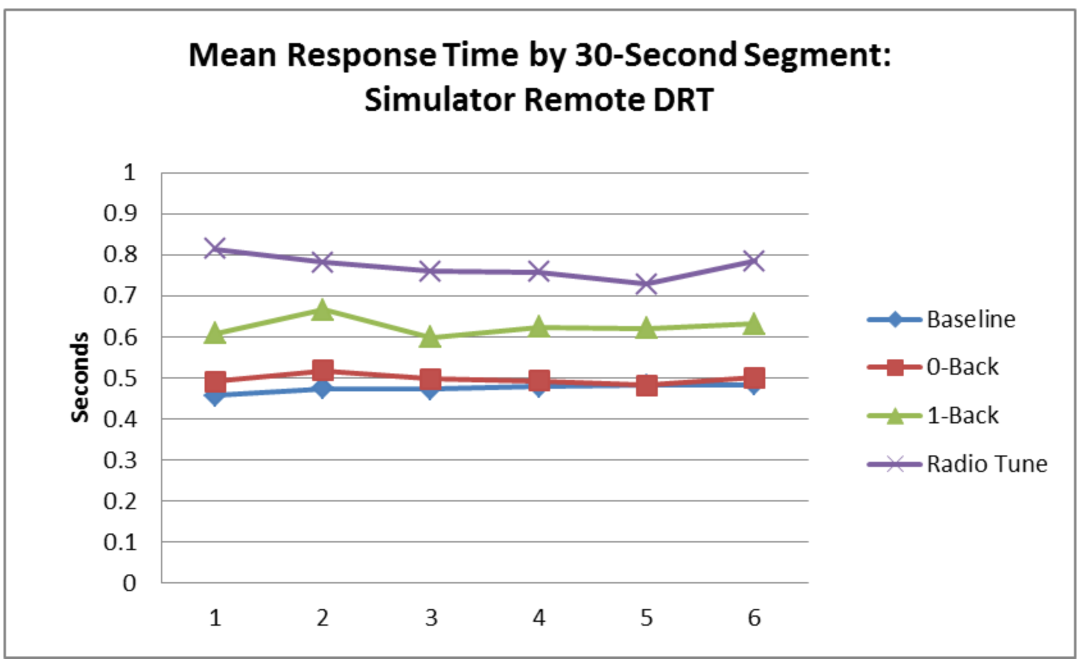


Figure 12. Mean Response Time by 30-Second Segment: Remote DRT in Driving Simulator

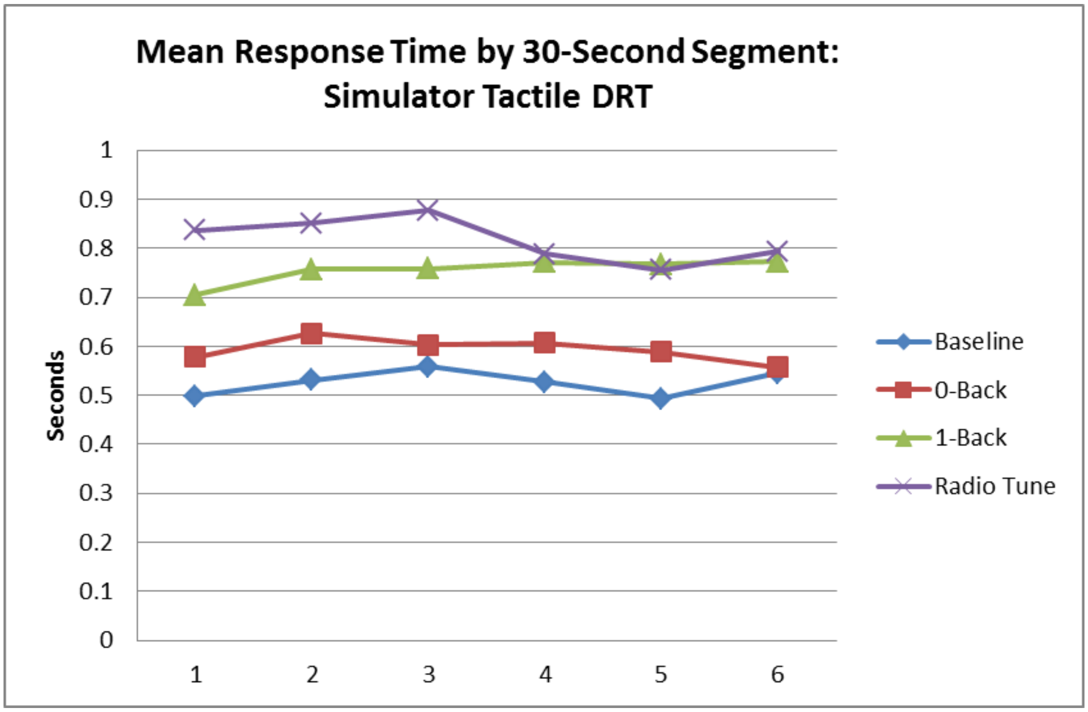


Figure 13. Mean Response Time by 30-Second Segment: Tactile DRT in Driving Simulator

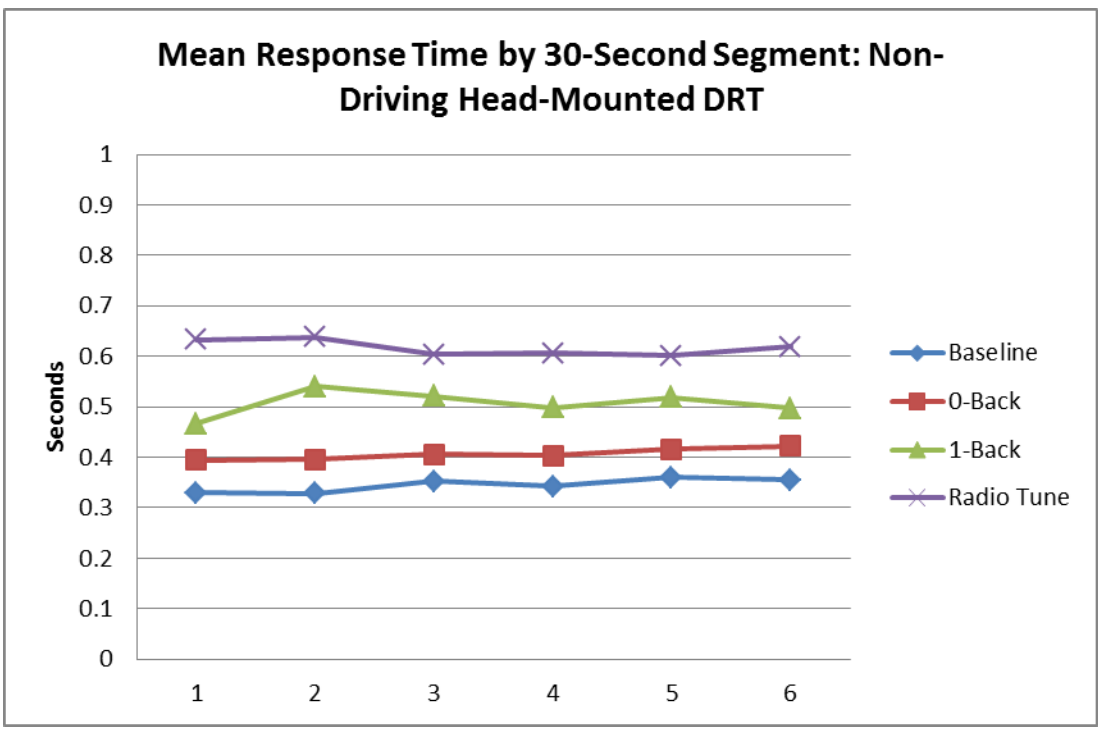


Figure 14. Mean Response Time by 30-Second Segment: Head-Mounted DRT Non-Driving

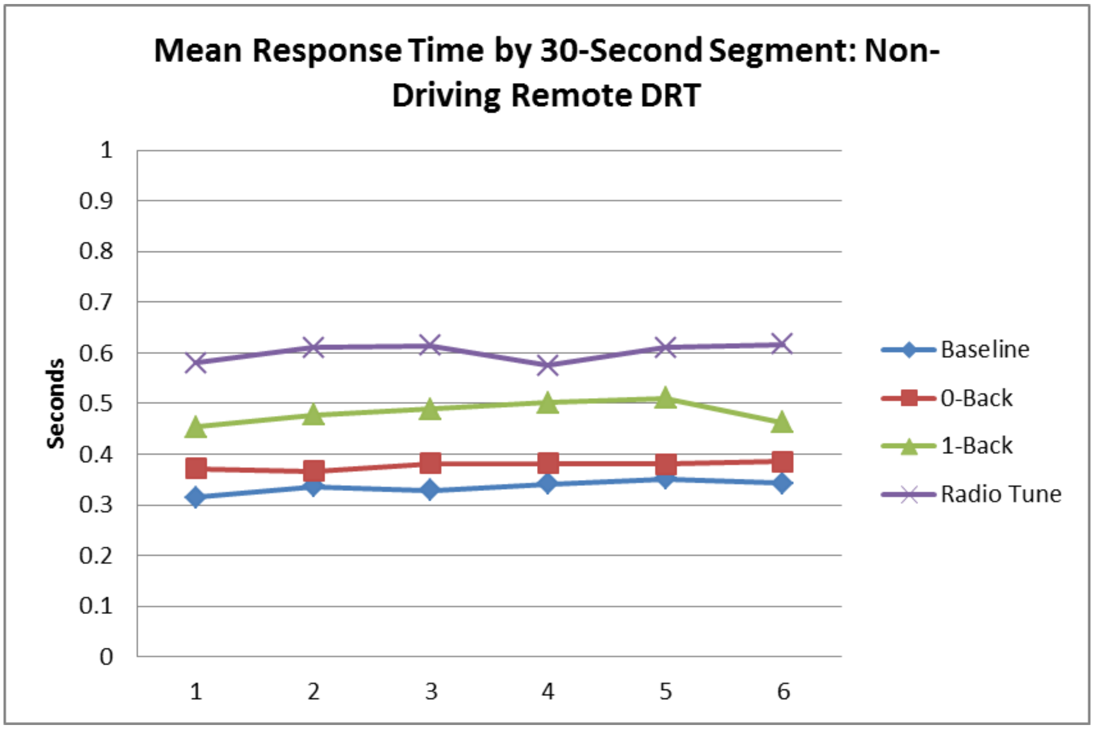


Figure 15. Mean Response Time by 30-Second Segment: Remote DRT Non-Driving

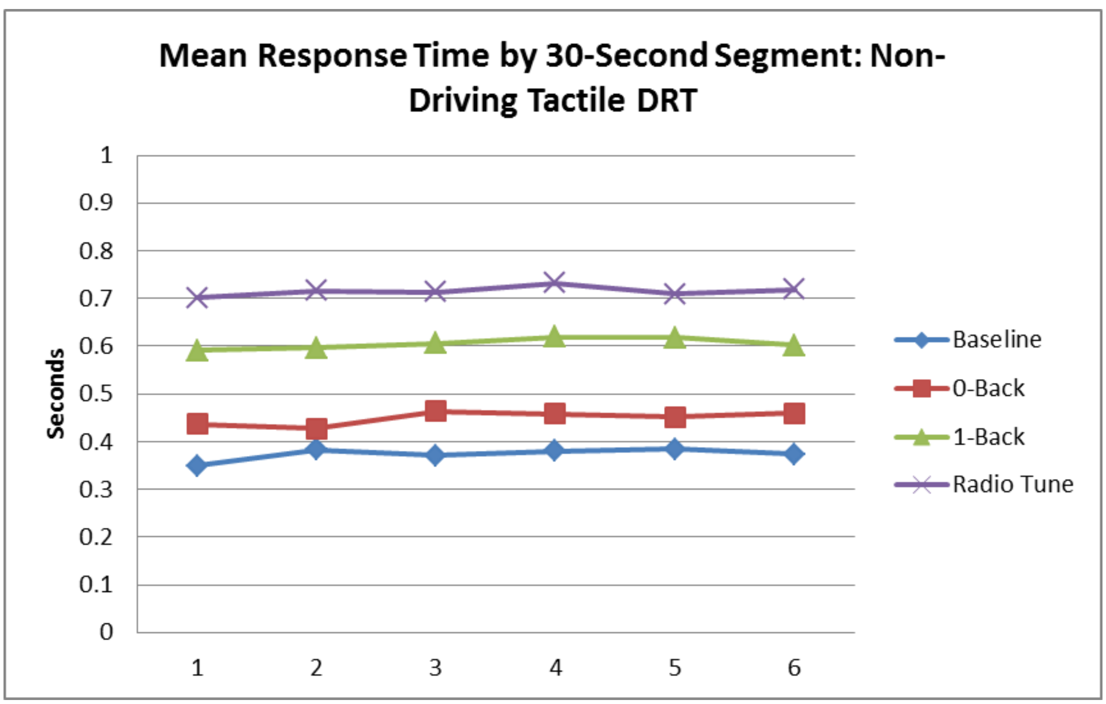


Figure 16. Mean Response Time by 30-Second Segment: Tactile DRT Non-Driving

The patterns of response time means across 30-second segments differ between the two test venues. Generally, the time sequences of response time means appear slightly more variable in the driving test conditions than in the non-driving test conditions. This difference is likely a direct reflection of the inherent differences between the test conditions, with increased variability due to the requirement to perform two tasks concurrently. The amount of separation among response time means associated with the secondary tasks appears somewhat smaller and less consistent across the 30-second segments in the driving conditions than in the non-driving conditions. This pattern is most apparent in the baseline and 0-back conditions. The absolute separation between these conditions is smallest in the RDRT simulator condition; the corresponding difference in the HDRT simulator condition is slightly greater but more variable across 30-second segments due to changes in the 0-back condition response time means. The difference between these two conditions is more consistent across 30-second segments in the TDRT simulator condition until the last 30-second segment, when the difference disappears.

Also apparent in the TDRT simulator condition is the change in relation between radio tuning and 1-back response time means over the course of the data collection interval. A relatively large and consistent difference apparent during the first three segments disappears in the fourth segment and remains absent for the remainder of the extended data collection interval. More generally, the pattern of 30-second response time means in the TDRT simulator condition suggests the operation of time-related differences, with slight slopes apparent among several of the conditions. In contrast, the response time means in the non-driving conditions appear less volatile across the 30-second segments with consistent and slightly larger differences between conditions, especially between the baseline and 0-back conditions. Noteworthy, given the apparent time-related changes in the TDRT simulator condition are the stability and absence of apparent time-related changes in the non-driving TDRT condition.

In consideration of the need to determine the minimum data collection interval necessary for obtaining valid data, the patterns of 30-second interval means suggest that the differences that appear initially remain fairly consistent over the first three 30-second segments, with some changes occurring after 90 seconds for the driving conditions. For the non-driving test conditions, the differences remain consistent across all six 30-second data segments.

Separate two-way analyses of variance (ANOVAs) were computed for each combination of test venue and DRT to provide a slightly more rigorous determination of whether there were systematic time-related changes. Secondary task Condition and Segment were the independent variables and Mean Response Time was the dependent variable. Analyses were computed over a 2-minute data collection interval to simplify interpretation, and because data collection intervals longer than two minutes are not currently under consideration by ISO. Probability values associated with F tests are presented in Table 4.

Table 4. Summary of Statistical Tests Using 30-Second Segments (Mean Response Time)

Effect	Simulator HDRT	Simulator RDRT	Simulator TDRT	Non-Driving HDRT	Non-Driving RDRT	Non-Driving TDRT
Condition	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001
Segment (30)	0.0432	0.3318	0.0935	0.3821	0.0842	0.3688
Condition* Segment	0.5871	0.5407	0.8024	0.2003	0.4653	0.9966

For these analyses, test results with probability values less than .05 are considered statistically significant, reflecting meaningful time-related changes; probability values between .05 and .10 are considered marginal, suggesting the potential for time-related changes. In Table 4, all of the Condition main effects were statistically significant, which indicates that all test venue by DRT combinations were successful in detecting at least one significant difference between secondary task conditions. These main effects will be examined in greater detail in later sections. The immediate concern is the existence of time-related effects, reflected by significant Segment main effects or by Condition x Segment interactions, which indicate that the differences between conditions do not remain consistent across the entire data collection interval. As shown in Table 4, none of the Condition x Segment interactions was significant, which indicates that the differences among the secondary task conditions were consistent across successive 30-second segments. One of the Segment main effects (Simulator HDRT) was statistically significant and two others (Simulator TDRT and Non-Driving RDRT) were marginal. The Simulator HDRT Segment main effect reflects a generalized increase in response time that occurred between the first and second 30-second segments, followed by a decrease between the second and third segments, which are apparent in Figure 11. Similar increases are apparent for the other two effects that approached statistical significance.

3.2 Response Time by 15-Second Segments

To accommodate the use of DRT methods with short tasks and also to reduce the need to combine data from multiple trials, the feasibility of using data collection intervals shorter than 30 seconds has been proposed by some researchers. In this section, we considered the stability of the response time metric across successive 15-second segments. With a presentation rate of 15 DRT targets per minute, each 15-second segment mean would be computed from at most 4 target presentations. Target detection response time means computed for successive 15-second segments for each combination of DRT variant and test venue are presented in Figure 17 through Figure 22.

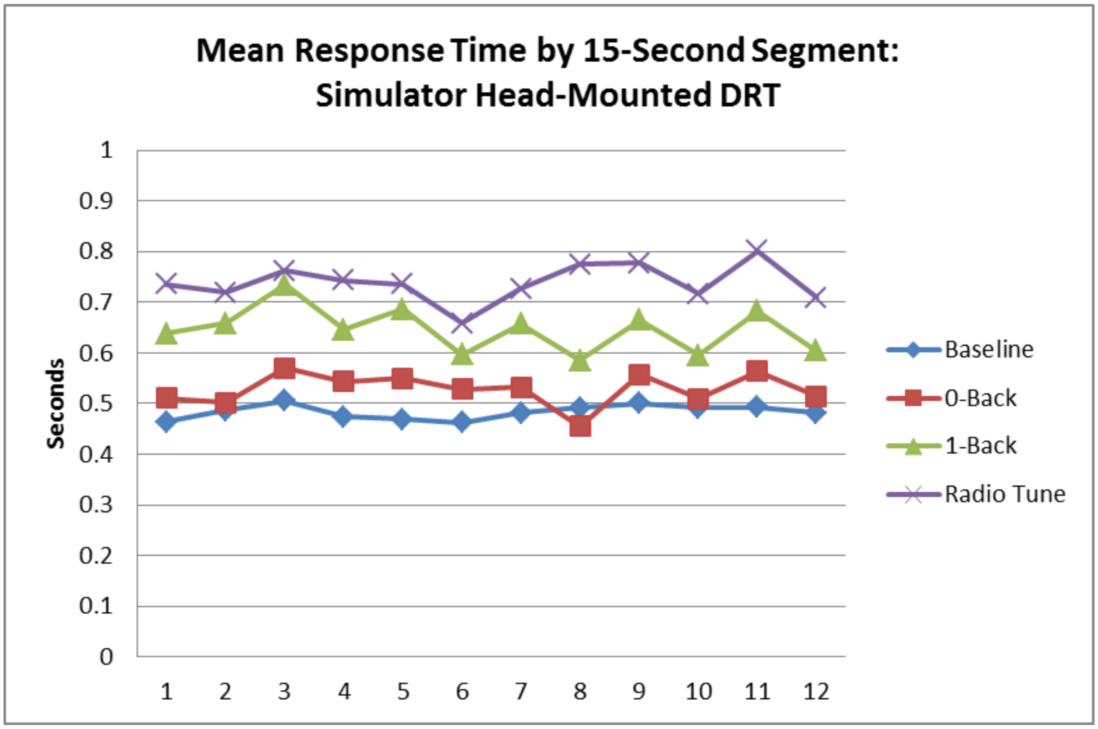


Figure 17. Mean Response Time by 15-Second Segment: Head-Mounted DRT Driving Simulator

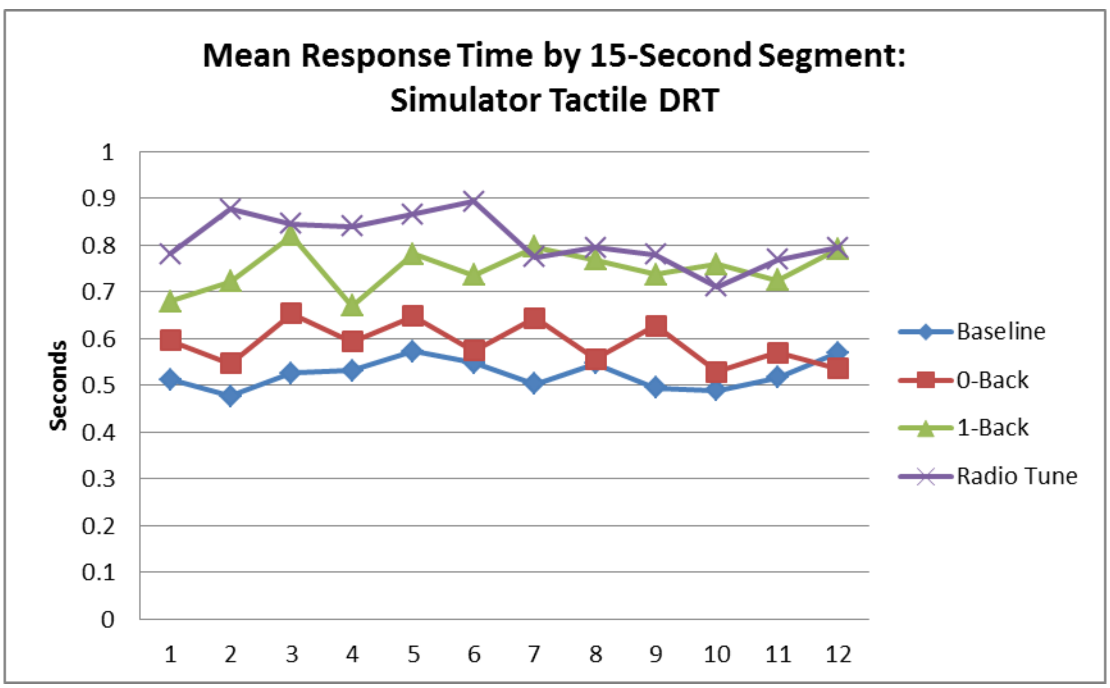


Figure 18. Mean Response Time by 15-Second Segment: Tactile DRT Driving Simulator

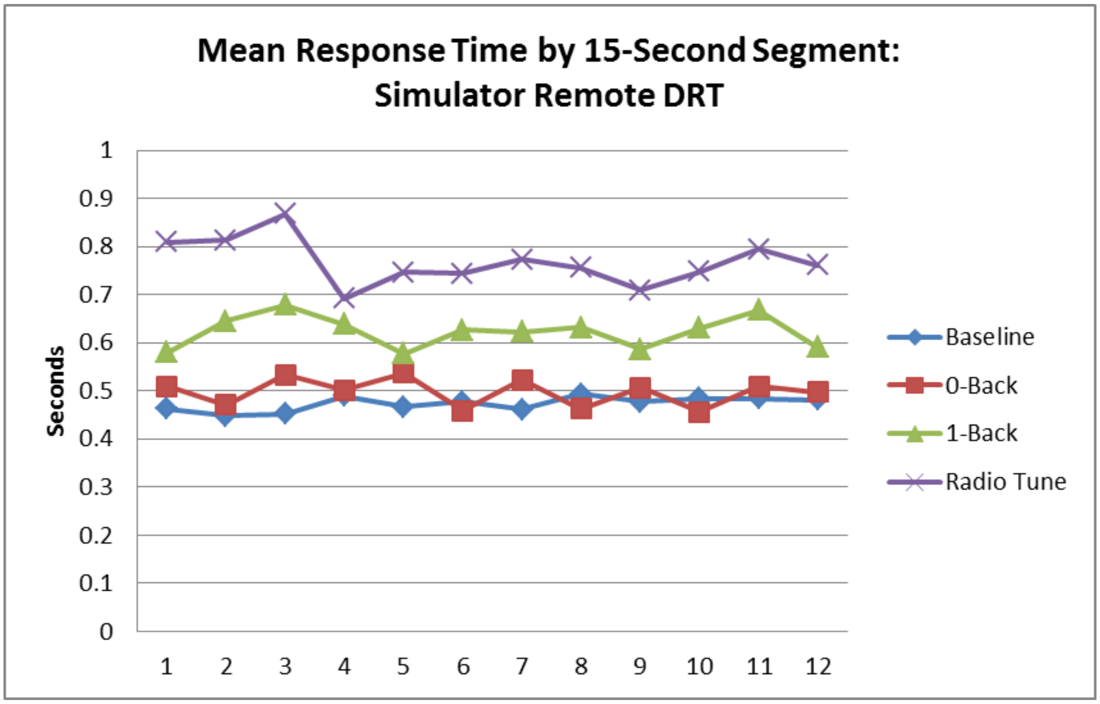


Figure 19. Mean Response Time by 15-Second Segment: Remote DRT Driving Simulator

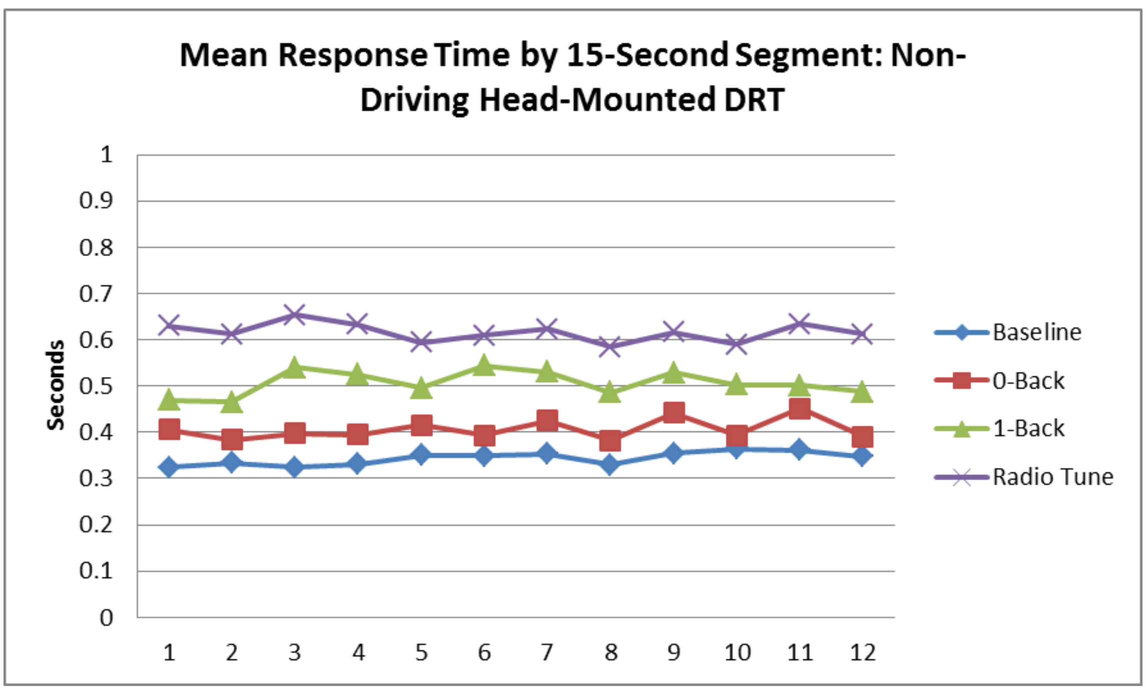


Figure 20. Mean Response Time by 15-Second Segment: Head-Mounted DRT Non-Driving

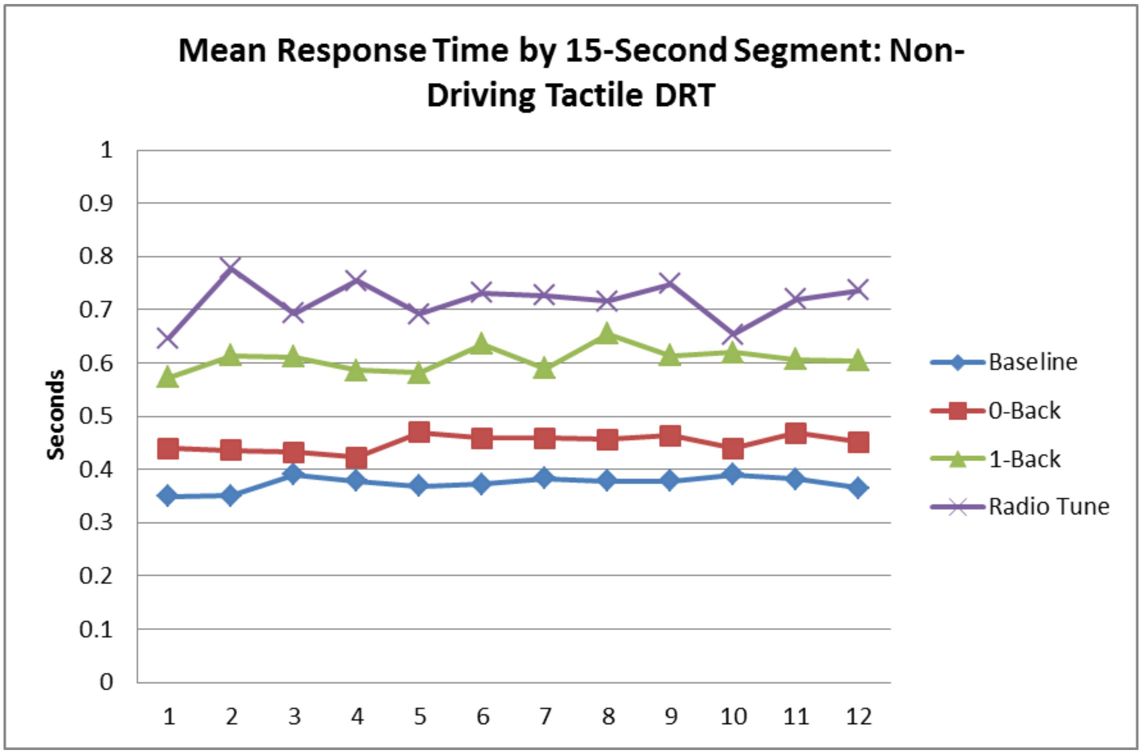


Figure 21. Mean Response Time by 15-Second Segment: Tactile DRT Non-Driving

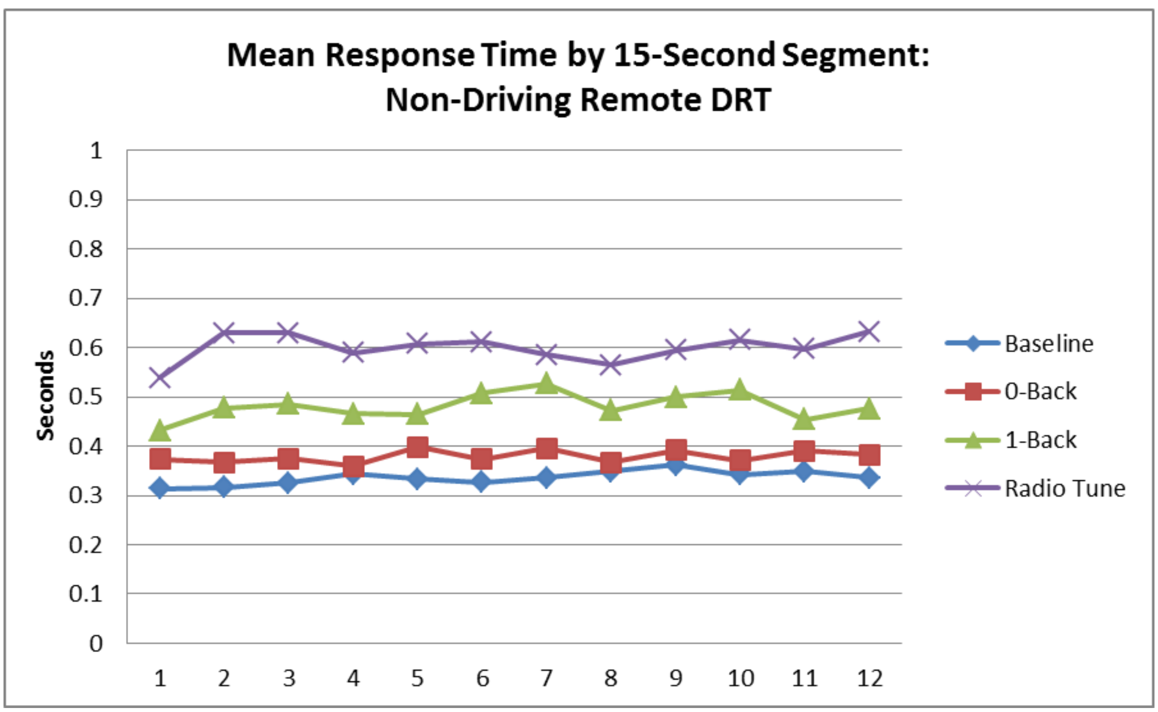


Figure 22. Mean Response Time by 15-Second Segment: Remote DRT Non-Driving

The sequences of means computed from 15-second segments exhibit a considerable amount of variability. Some of the time traces, particularly those in the driving simulator conditions, are characterized by saw tooth patterns, in which mean values alternate between higher and lower values. As was evident among the 30-second traces, more variability is apparent in the driving simulator conditions than in the non-driving conditions. The differences among the four traces also remain relatively consistent across the entire data collection interval in the non-driving conditions. In the driving simulator conditions, the relationship among the four traces does not remain consistent across the entire data collection interval. The existence of marked changes in the response time means between the first and second 15-second segments is apparent in several of the test venue by DRT combinations, even among the non-driving conditions. This pattern suggests the existence of generalized start-up effects, which if real, would imply that a single 15-second interval would not be sufficient to provide stable estimates of DRT response times associated with the different secondary task conditions.

ANOVAs were computed for each combination of test venue and DRT to examine the time-related trends more carefully. The probability values associated with F tests are presented in Table 5.

Table 5. Summary of Statistical Tests Using 15-Second Segments (Mean Response Time)

Effect	**Simulator HDRT**	**Simulator RDRT**	**Simulator TDRT**	**Non-Driving HDRT**	**Non-Driving RDRT**	**Non-Driving TDRT**
Condition	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001
Segment (15)	0.0062	0.1370	0.0652	0.1165	0.0165	0.2929
Condition* Segment	0.4687	0.0306	0.4145	0.6936	0.1892	0.7113

Two of the six Segment main effects were statistically significant (Simulator HDRT; Non-Driving RDRT) and a third one was marginal (Simulator TDRT). These significant main effects indicate that mean response times varied systematically during the data collection interval. Because the main effect represents data collapsed across all secondary task conditions, it is difficult to interpret the trend directly; rather, the significant main effect should be taken as an indication that the response time measures in these conditions were less consistent across the 2-minute data collection interval than in other test venue by DRT combinations.

The significant Condition X Segment interaction in the Simulator RDRT combination is more amenable to direct interpretation, indicating that the differences among some secondary task conditions are not consistent across the entire data collection interval. As shown in Figure 19, which represents this effect, several of the response time means exhibit a high level of volatility in the early segments (1-4). Specifically, the difference between mean response times in the radio-tuning and 1-back conditions changes dramatically between the first and fourth 15-second segments. Similarly, the difference between the 1-back and 0-back condition mean response times exhibits a large change over the first five segments. The statistical significance of this interaction represents a potentially more serious problem for the metrics over time than the significant Segment main effect because the interaction indicates that comparisons made using different segments of a longer data collection interval could lead to different conclusions. Overall however, the statistical test results indicate that differences between secondary task conditions are generally stable over a 2-minute data collection interval when the full sample (N = 48) is used as the basis for analysis. The fact that one of the six interactions was statistically significant when 15-second segments were used versus none of the six when 30-second segments were used reflects slightly greater stability of the response time metric in the 30-second segments. These findings reflect the stability of the metrics and consistency of the differences between conditions when analyses focus on different parts of a longer data collection interval. They do not have direct implications for determining the optimal interval duration that is appropriate for testing, other than to suggest that a 15-second data collection interval probably is too volatile to provide stability in the response time metric. Analyses presented later will compare data collection intervals of different durations using a family of planned comparisons.

3.3 Hit Rate by 30-Second Segments

Mean hit rate values, computed using 30-second segments, are presented in Figure 23 through Figure 28 for each combination of DRT condition and test venue.

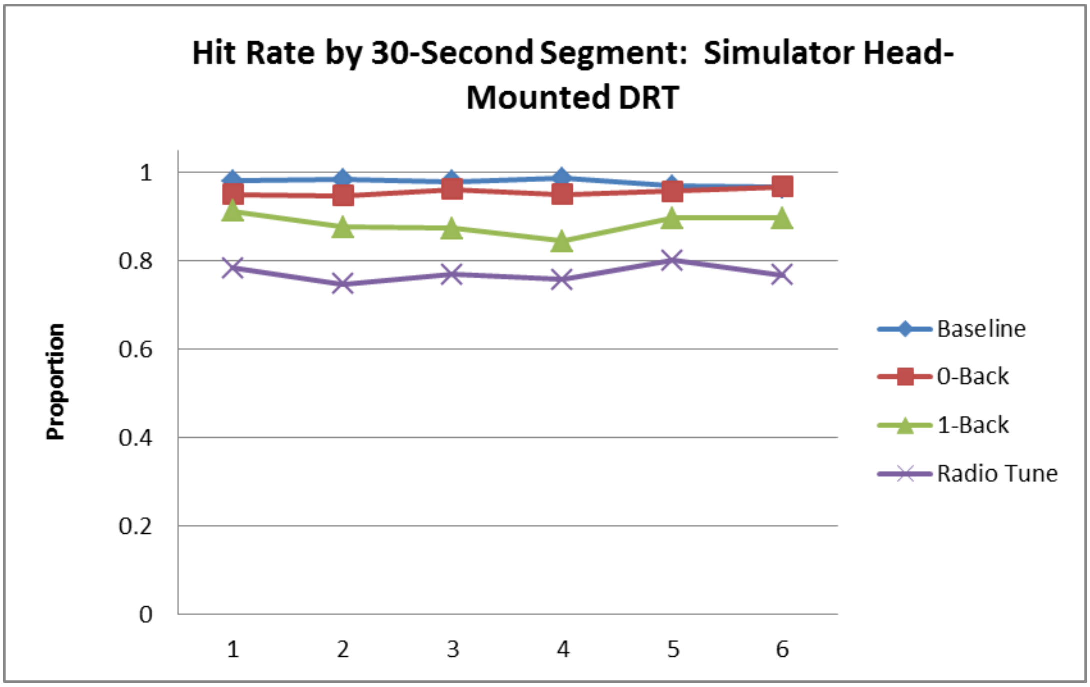


Figure 23. Mean Hit Rate by 30-Second Segment: Head-mounted DRT in Driving Simulator

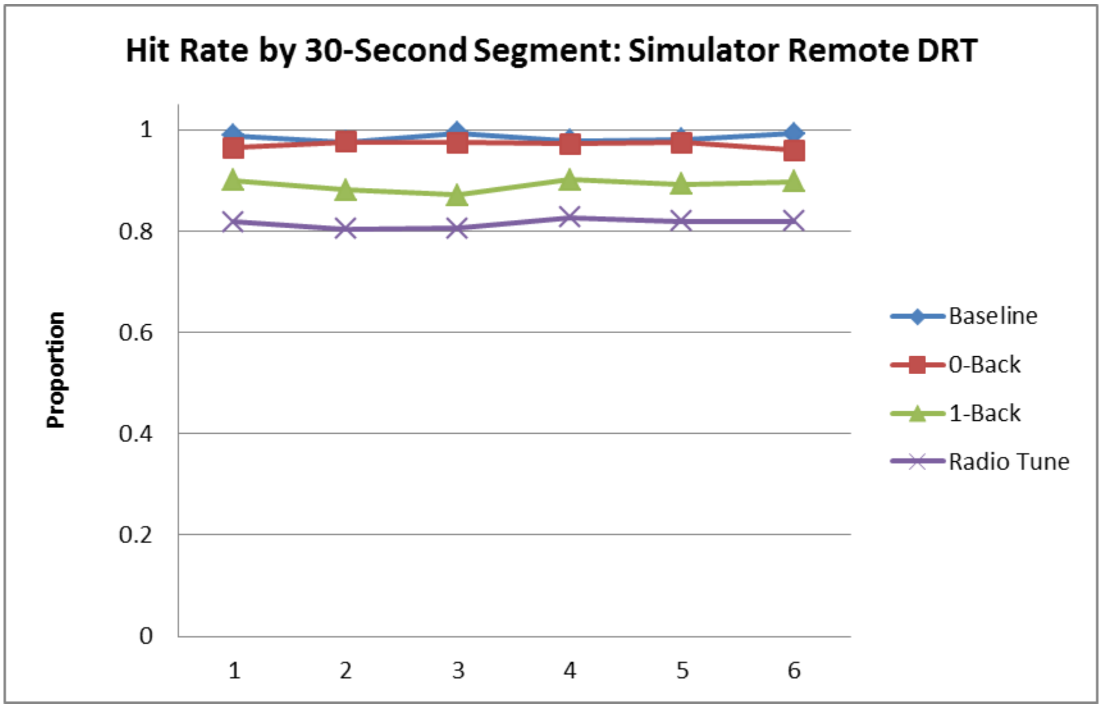


Figure 24. Mean Hit Rate by 30-Second Segment: Remote DRT in Driving Simulator

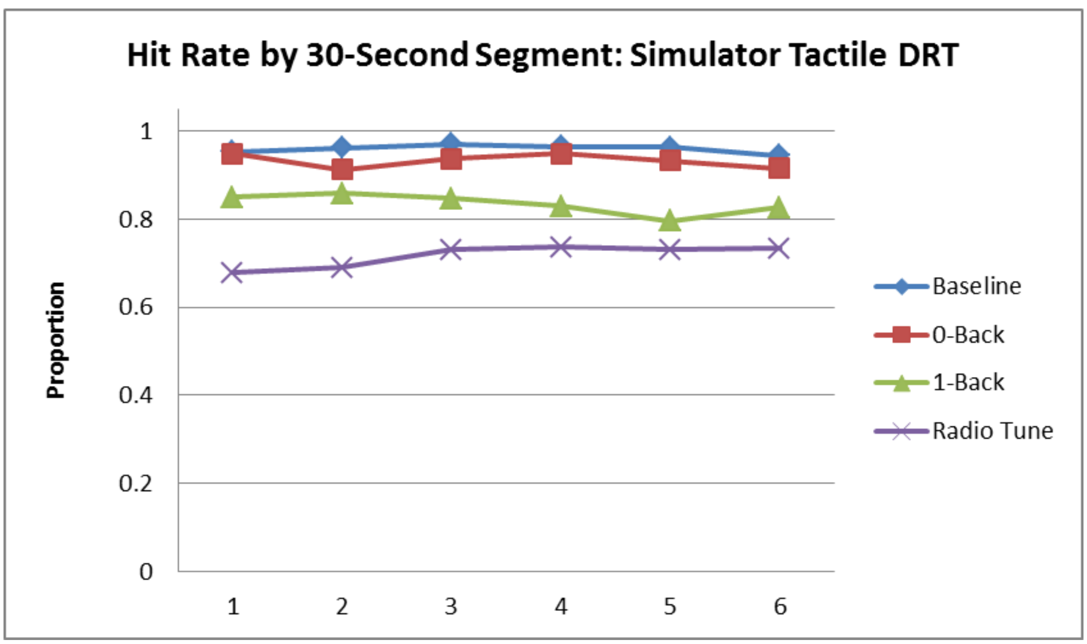


Figure 25. Mean Hit Rate by 30-Second Segment: Tactile DRT in Driving Simulator

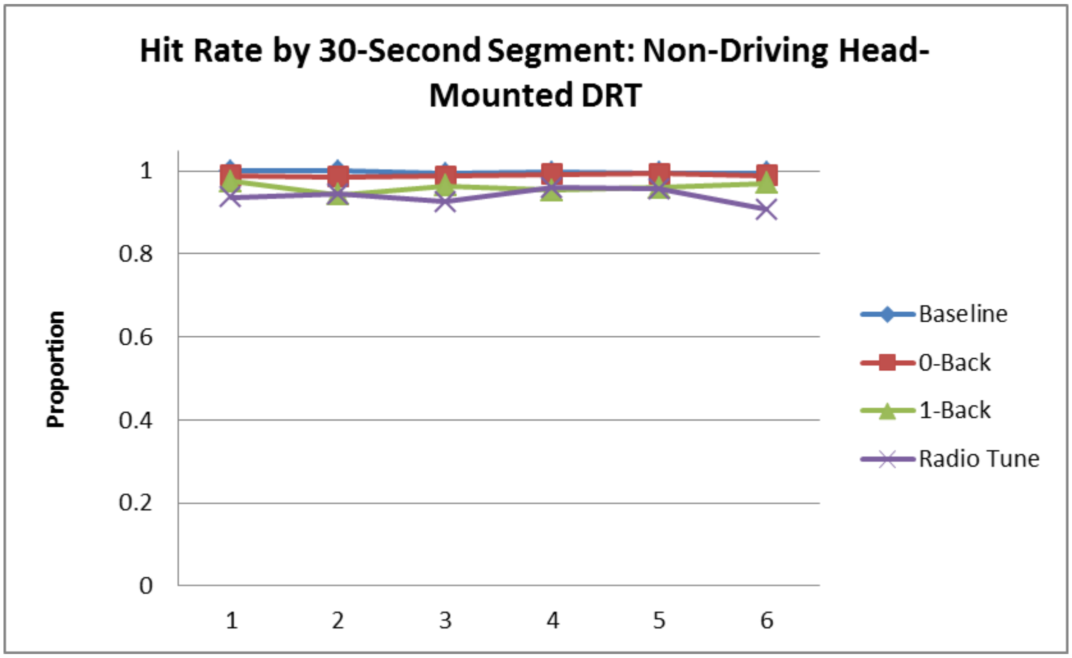


Figure 26. Mean Hit Rate by 30-Second Segment: Head-mounted DRT Non-Driving

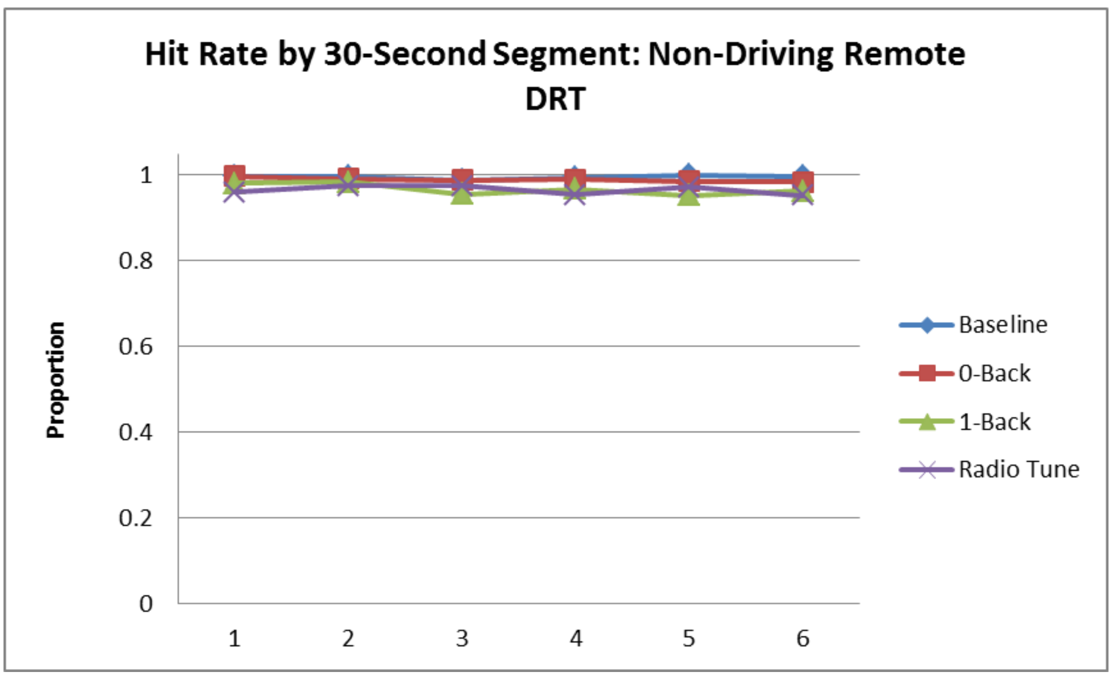


Figure 27. Mean Hit Rate by 30-Second Segment: Remote DRT Non-Driving

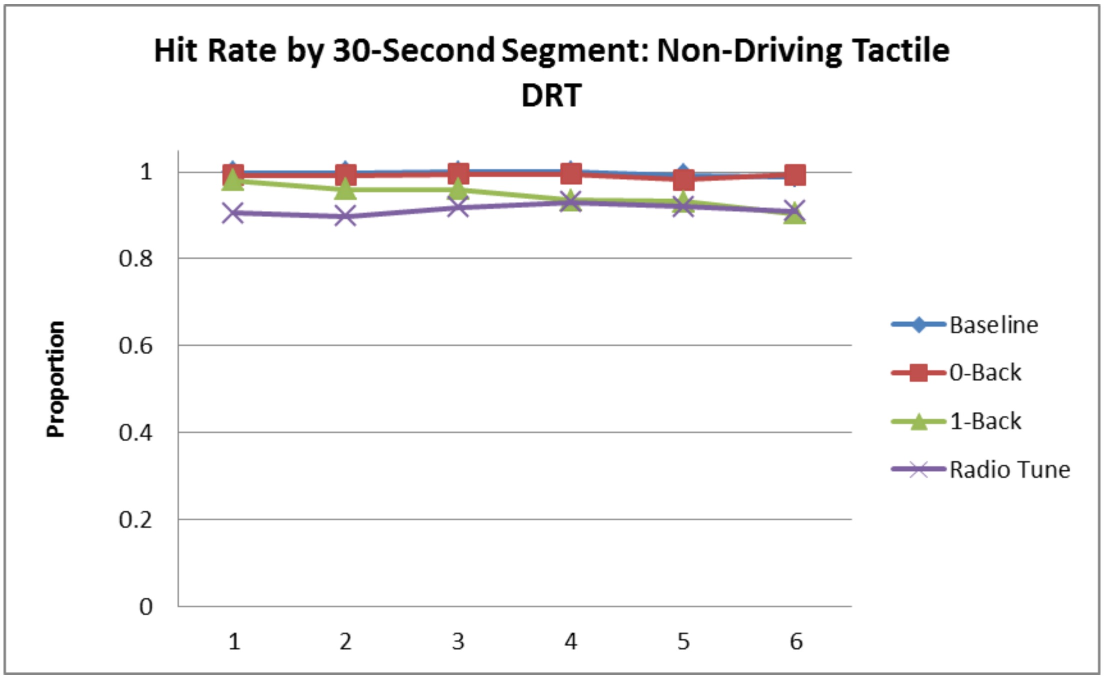


Figure 28. Mean Hit Rate by 30-Second Segment: Tactile DRT Non-Driving

Overall, the mean hit rate values appear somewhat less variable over time than the response time means. As noted earlier, the simulator venue hit rates exhibited generally larger differences among the secondary task conditions than the non-driving venue. In the simulator venue, the differences between the baseline and 0-back conditions were consistently smaller than the other differences. Differences among the three secondary task conditions were apparent in all three simulator DRT conditions, with the largest differences evident for the TDRT simulator condition. The TDRT means in the simulator venue were also associated with more variability over time than the other two DRT variants in the simulator. Differences among the secondary task conditions in hit rates were much less apparent in the non-driving venue, due to the near-perfect performance in this test venue. Among the non-driving DRT conditions, the TDRT reveals greater separation among the hit rates associated with the more difficult tasks, however the differences did not remain consistent over time.

ANOVAs were computed using hit rates to identify trends over time. Probability values associated with the F tests are presented in Table 6.

Table 6. Summary of Statistical Tests Using 30-Second Segments (Mean Hit Rate)

Effect	**Simulator HDRT**	**Simulator RDRT**	**Simulator TDRT**	**Non-Driving HDRT**	**Non-Driving RDRT**	**Non-Driving TDRT**
Condition	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001
Segment (30)	0.3921	0.8214	0.6180	0.5628	0.1648	0.8326
Condition* Segment	0.7932	0.9583	0.6247	0.4138	0.4032	0.2629

None of the Segment main effects and none of the Condition x Segment interaction effects were statistically significant. This outcome is a bit surprising in consideration of the Non-Driving TDRT means, shown in Figure 28, which appear to suggest the existence of a Condition x Segment interaction. Time-related changes appear to exist for both the Radio Tuning and 1-back conditions. However as shown in Table 6, this effect was not close to being statistically significant. More generally, the absence of any statistically significant time-related changes reflects the greater stability of the hit rate metric over time, based on the 30-second data collection segments.

3.4 Hit Rate by 15-Second Segments

The hit rate means were also computed using 15-second intervals. The time traces for these means are presented for each combination of test venue by DRT condition in Figure 29 through Figure 34.

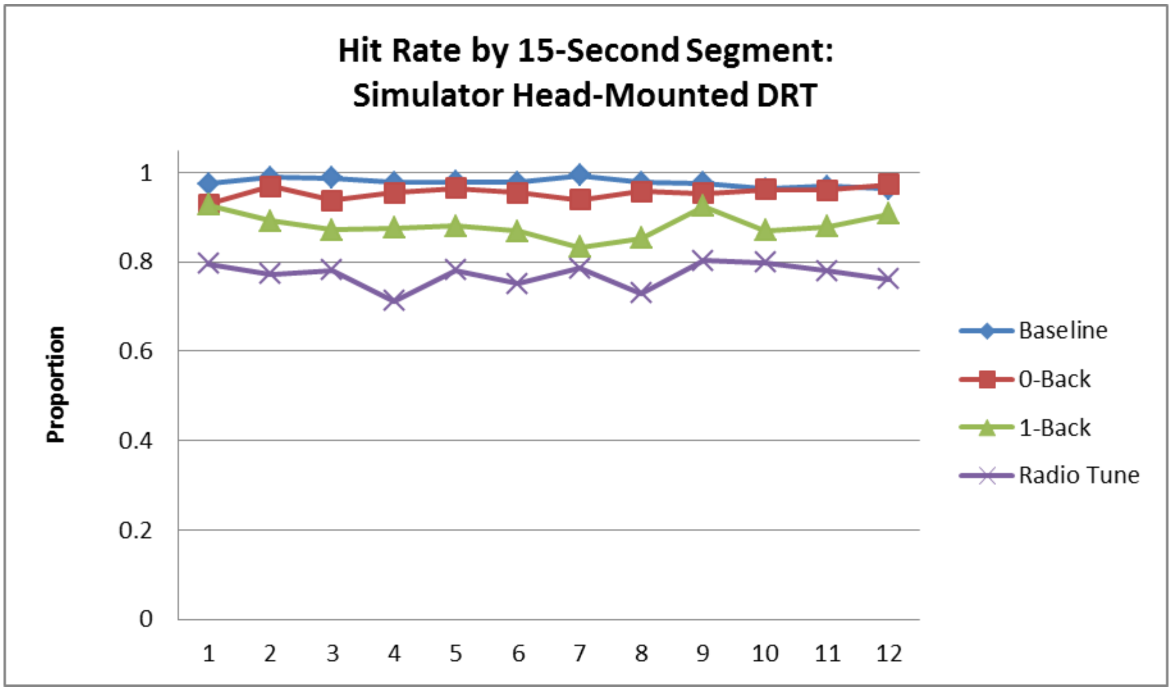


Figure 29. Mean Hit Rate by 15-Second Segment: Head-Mounted DRT in Driving Simulator

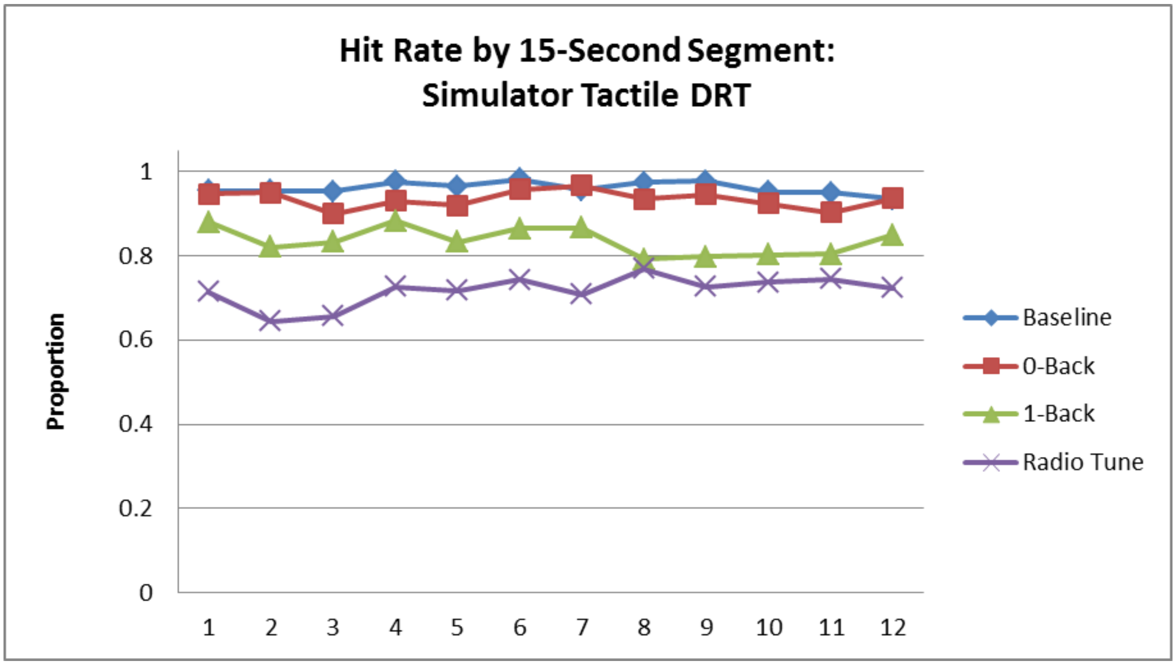


Figure 30. Mean Hit Rate by 15-Second Segment: Tactile DRT in Driving Simulator

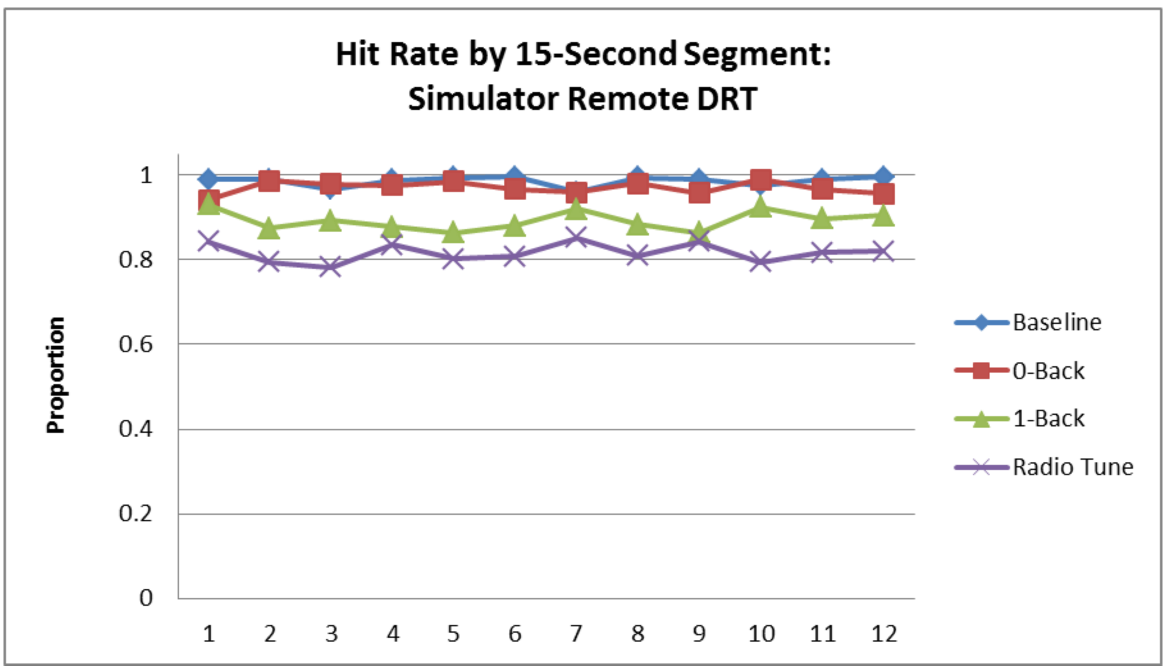


Figure 31. Mean Hit Rate by 15-Second Segment: Remote DRT in Driving Simulator

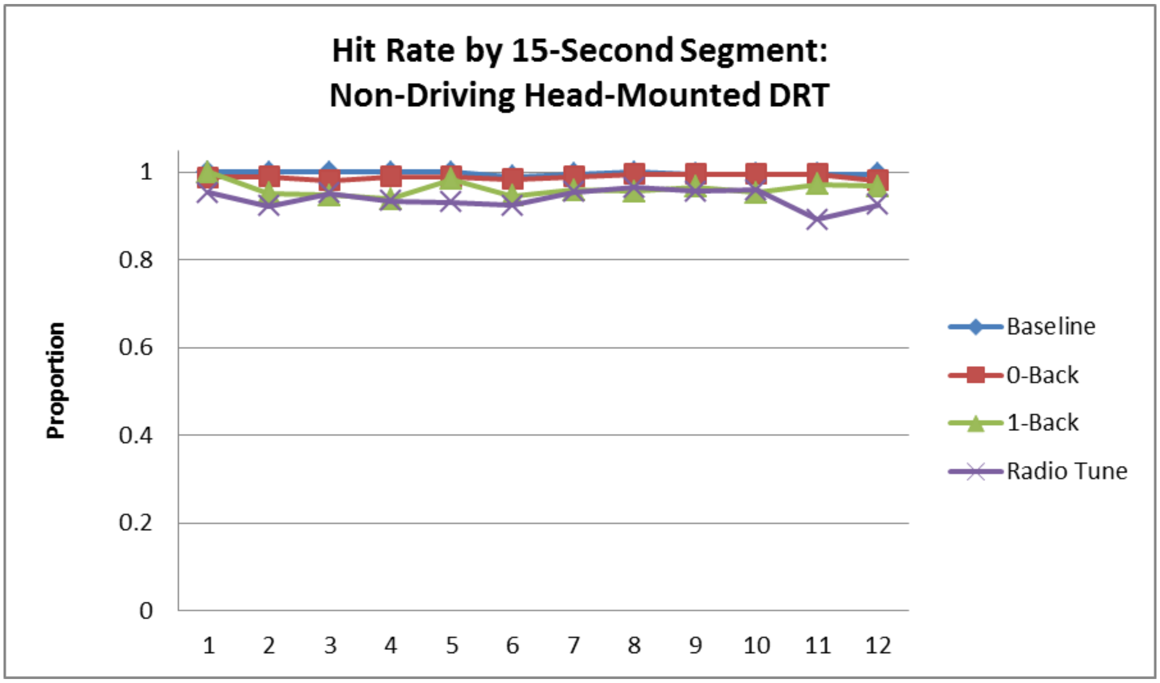


Figure 32. Mean Hit Rate by 15-Second Segment: Head-Mounted DRT Non-Driving

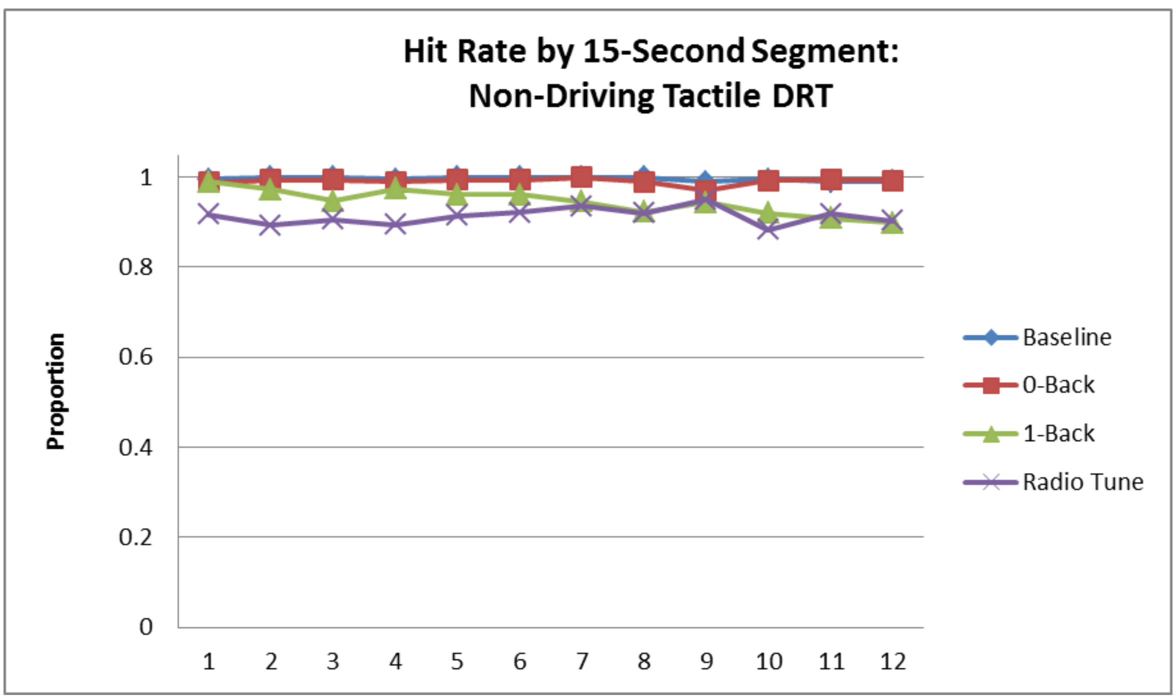


Figure 33. Mean Hit Rate by 15-Second Segment: Tactile DRT Non-Driving

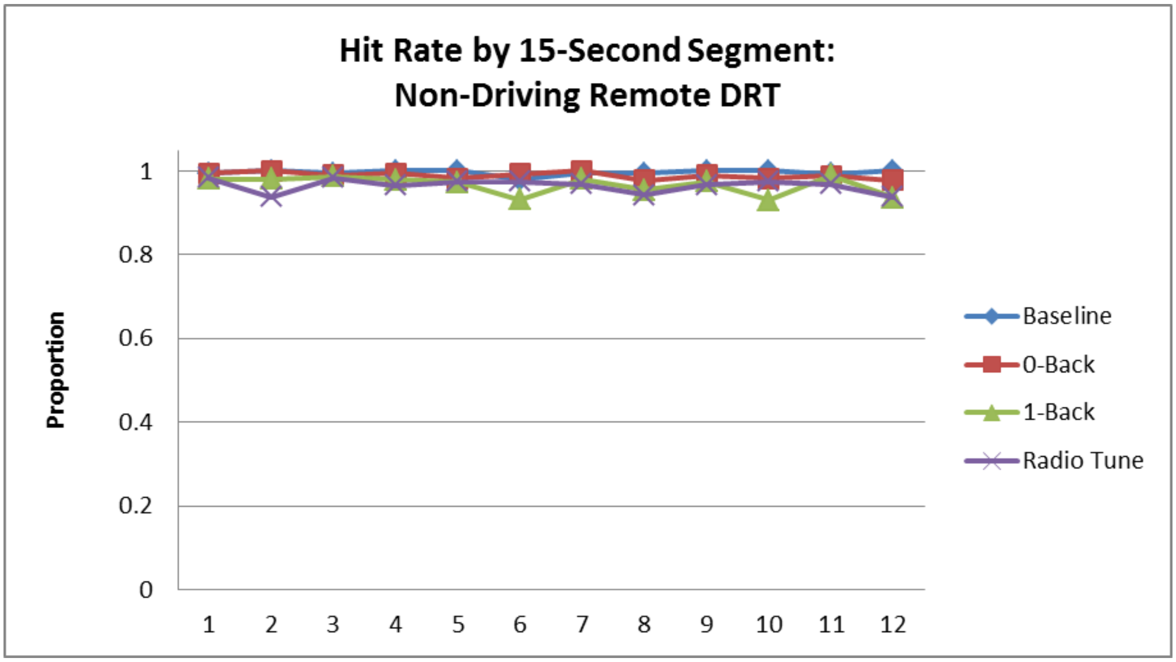


Figure 34. Mean Hit Rate by 15-Second Segment: Remote DRT Non-Driving

The 15-second DRT hit rate means generally exhibit more variability in the driving simulator venue than in the non-driving venue. Approximations of the saw tooth pattern are apparent in some traces. A summary of statistical test results is presented in Table 7.

Table 7. Summary of Statistical Tests Using 15-Second Segments (Mean Hit Rate)

Effect	**Simulator HDRT**	**Simulator RDRT**	**Simulator TDRT**	**Non-Driving HDRT**	**Non-Driving RDRT**	**Non-Driving TDRT**
Condition	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001
Segment (15)	0.5659	0.9093	0.0791	0.1955	0.0303	0.8833
Condition*Segment	0.7007	0.5464	0.4948	0.6406	0.1354	0.5910

One of the Segment main effects (Non-Driving RDRT) was statistically significant and another (Simulator TDRT) approached statistical significance; these effects indicate a systematic time-related change in hit rate across all secondary task conditions. The first effect, which represents a combination of the traces shown in Figure 34, is difficult to interpret given the lack of separation among the various secondary task conditions; however, the aggregated hit rate means for this condition reveal a slight downward trend over the first eight time segments. The second effect, aggregated from the traces shown in Figure 30, is difficult to discern; the improvement in detection task performance over time in the radio tuning condition appears unique to that secondary task condition and is thus more consistent with an interaction effect than a main effect.

3.5 Differences Between DRT by Test Venue Combinations

Statistical analyses were performed to determine: (1) whether there are differences among the various DRT/test venue combinations in their sensitivity for differentiating between tasks with known differences in attentional demand; (2) how the DRT performance degradation associated with visual-manual radio tuning compares with that associated with the 0-back and 1-back conditions; and (3) what is the minimum data collection interval necessary to obtain meaningful differences for the purposes of NHTSA Guidelines testing. These objectives are embodied in the following specific questions.

1. Do differences exist among DRT conditions (HDRT, TDRT and RDRT) that would make one preferable for use in testing?
2. Do differences exist between test venues (Simulator, Non-Driving) that would make one preferable for use in testing?
3. Do differences exist among data collection intervals of different durations?

These questions were addressed (1) with a set of omnibus ANOVAs, in which data from all conditions were combined; and (2) by assessing the outcomes of a set of planned comparisons, which compared the sensitivity of the various test venue by DRT combinations directly; (3) by assessing test-retest reliability; and (4) by comparison of effect sizes (ESs).

Three-factor ANOVAs were computed with DRT, Venue, and Condition as the independent variables. Mean Response Time was the dependent variable. The summary of these analyses is presented in Table 8, in which F is the ANOVA F-test statistic, and p is the probability value.

Table 8. ANOVA Summary Tables by Data Collection Interval: Mean Response Time

	2 Minute		1 Minute		30 Seconds	
Effect	**F**	**p**	**F**	**p**	**F**	**p**
DRT	61.88	<.0001	34.44	<.0001	23.96	<.0001
Venue	422.44	<.0001	333.51	<.0001	236.6	<.0001
DRT*Venue	0.46	0.6321	0.44	0.6451	0.95	0.3870
Condition	369.21	<.0001	277.64	<.0001	219.24	<.0001
DRT*Condition	2.37	0.0277	1.29	0.2598	1.3	0.2535
Venue*Condition	0.2	0.8995	0.16	0.9239	0.88	0.4521
DRT*Venue*Condition	0.78	0.5859	0.82	0.5572	1.36	0.2286

Statistical test results were generally consistent across the three time intervals with one exception. The marginally significant DRT x Condition interaction at the 2-minute data collection interval reflects small and generally inconsequential differences in the magnitude of differences between secondary task conditions in the different DRT venues. Much stronger were the three main effects, which were statistically significant across all three data collection intervals. The means for the Test Venue main effect are presented in Figure 35.

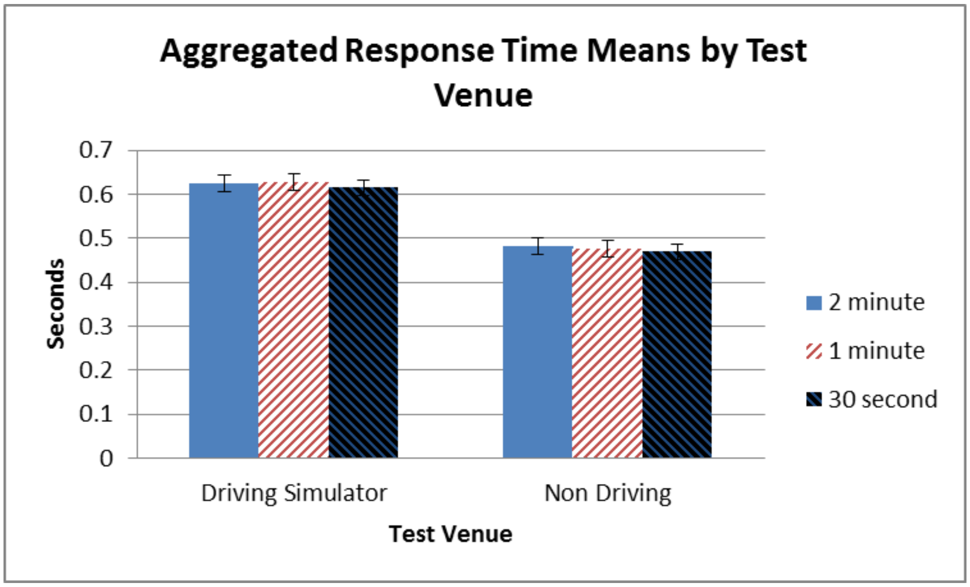


Figure 35. Aggregated Response Time Means by Test Venue and Data Collection Interval

As shown earlier, the simulator response time means were consistently slower than the non-driving venue response time means. The response time means associated with the DRT main effect are presented in Figure 36.

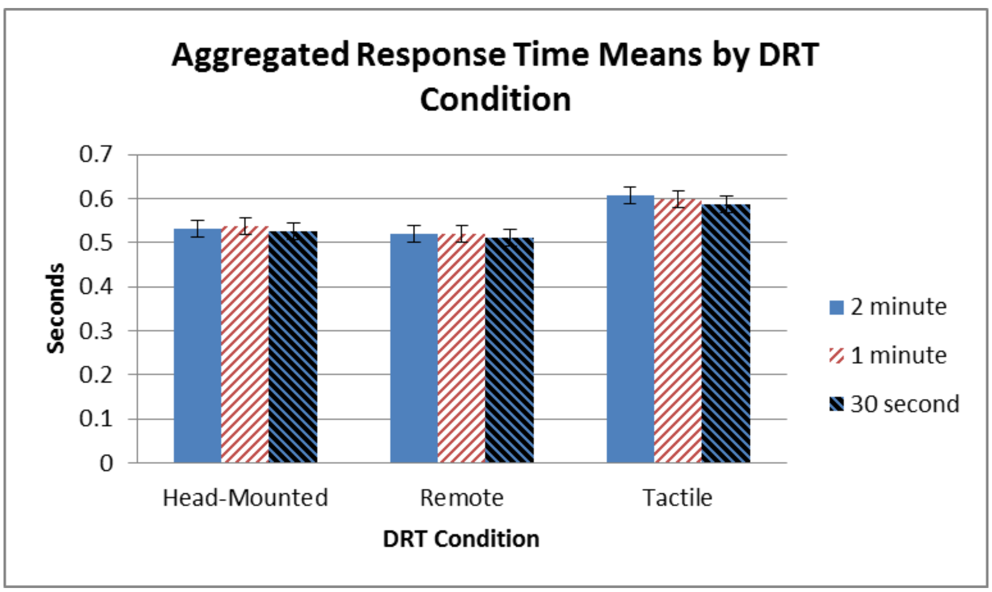


Figure 36. Aggregated Response Time Means by DRT and Data Collection Interval

Paired t-tests were performed on the DRT condition response time means for each time interval. The Tactile DRT response time means were significantly longer than those associated with the other two DRT conditions; these pairwise differences were consistent across the three data collection intervals. The differences between the other DRT conditions were not statistically significant.

Both sets of means exhibit a trend toward slightly faster response times in the shorter data collection intervals; however, it is not possible to test differences between the data collection intervals because they are overlapping intervals and not independent.

Based on the consistency of statistical test outcomes across the three data collection intervals, it appears that any of the three data collection intervals would provide identical results. However, this conclusion does not apply directly to the testing situation in which a single DRT by test venue combination is used to identify differences between secondary task conditions. The present results benefit from the fact that there were a relatively large number of participants in the data sample (N = 48) and from the fact that effects tested in these analyses used data collapsed over multiple conditions. These two factors combine to provide a relatively high level of statistical power, which is considerably greater than would be available in a small-sample test using a single DRT condition. The same three ANOVAs were computed for the DRT hit rate metric. A summary of the statistical test is presented in Table 9.

Table 9. ANOVA Summary Tables by Data Collection Interval: Hit Rate

	2 Minute		**1 Minute**		**30 Seconds**	
Effect	**F**	**p**	**F**	**p**	**F**	**p**
DRT	13.31	<.0001	15.55	<.0001	11.61	<.0001
Venue	248.54	<.0001	222.54	<.0001	151.03	<.0001
DRT*Venue	3.75	0.0238	3.93	0.0200	4.69	0.0093
Condition	140.61	<.0001	124.44	<.0001	89.58	<.0001
DRT*Condition	3.12	0.0049	4.04	0.0005	3.4	0.0025
Venue*Condition	47.25	<.0001	39.72	<.0001	26.28	<.0001
DRT*Venue*Condition	0.15	0.9893	0.4	0.8778	0.72	0.6346

The statistical test outcomes are consistent across the three data collection intervals. All main effects are statistically significant in each of the data collection intervals. However, unlike the pattern of results for mean response time, all of the two-way interactions are statistically significant for hit rate, which indicates that the pattern of findings is not consistent across all combinations of DRT, Venue and Secondary Task Condition. Examination of the results of the three analyses revealed that the interaction effects are virtually identical across the three data collection intervals. The DRT x Test Venue interaction effect for the 2-minute data collection interval is presented in Figure 37.

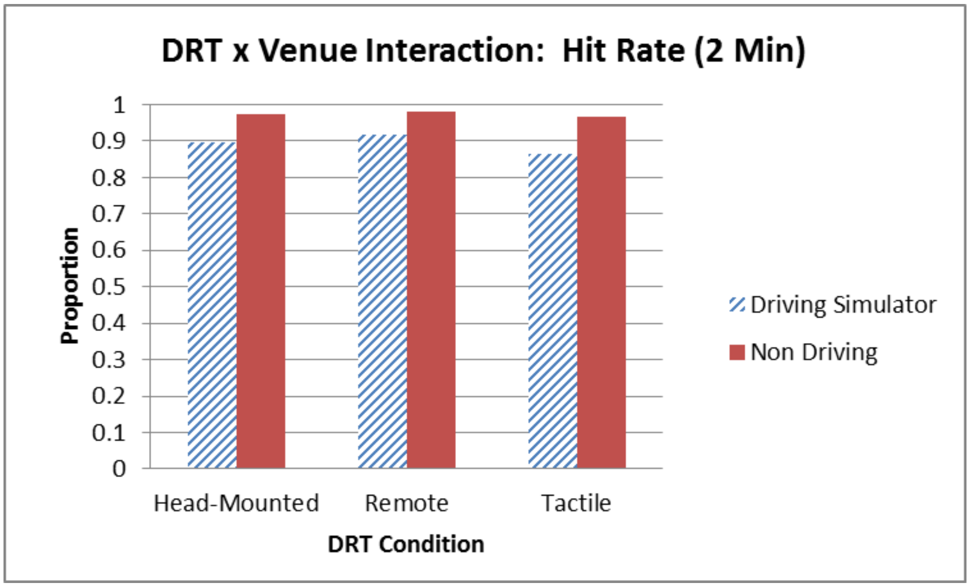


Figure 37. DRT x Test Venue Interaction: Hit Rate (2-Minute Data Collection Interval)

As shown, the mean hit rate values were consistently lower in the driving simulator venue than in the non-driving venue, however, the differences in hit rate among the various DRT conditions are greater in the driving simulator venue than in the non-driving venue. Most generally, the post hoc pairwise comparison testing revealed that hit rates in the DRT conditions were statistically different from one another in the driving simulator venue but not in the non-driving venue. The DRT x Condition interaction using data collapsed over test venue is presented in Figure 38.

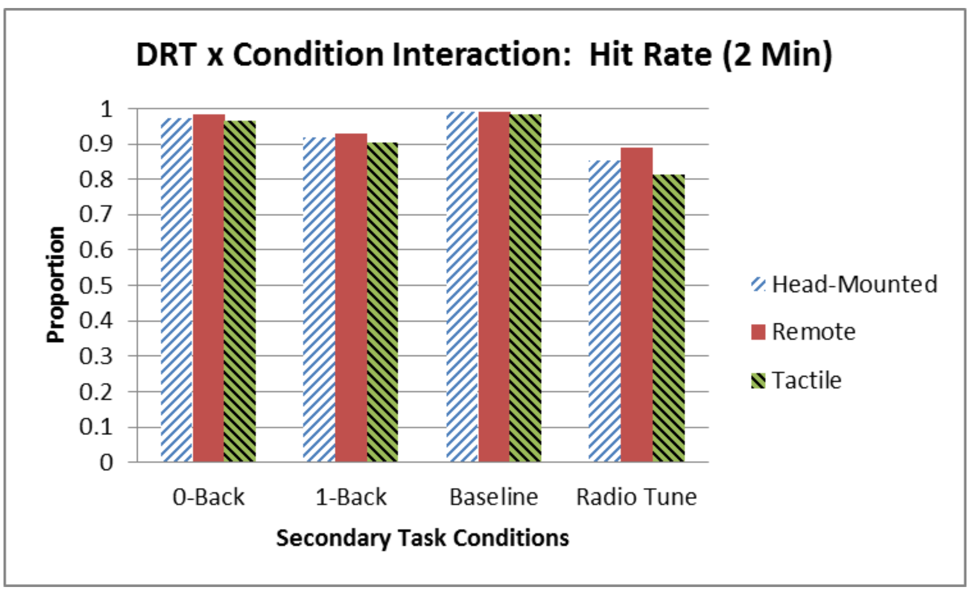


Figure 38. DRT x Condition Interaction: Hit Rate (2-Minute Data Collection Interval)

The essence of this interaction is the fact that there were no differences in hit rate among the three DRT conditions for any of the secondary task conditions except radio tuning, for which the hit rates were statistically different in each of the three DRT conditions; where TDRT had the lowest hit rates in this secondary task condition. The significant Venue x Condition interaction effect using data collapsed over DRT condition is presented in Figure 39.

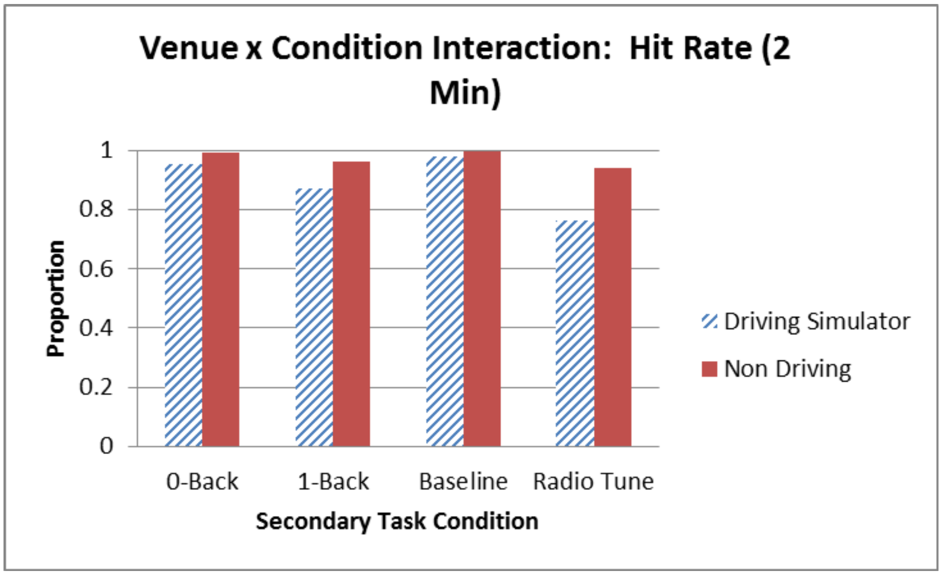


Figure 39. Venue x Condition Interaction: Hit Rate (2-Minute Data Collection Interval)

As shown, the differences in hit rates among secondary task conditions were smaller in the non-driving venue relative to those observed in the driving simulator venue. This effect, as well as the DRT x Test Venue interaction, reflects the ceiling effect due to the consistently high hit rate values observed in the non-driving venue. Most generally, the significant interaction effects indicate that hit rate has reduced sensitivity in the non-driving venue relative to the driving simulator venue.

3.6 Planned Comparison Analyses

The following four planned comparisons were used to examine the differences between DRT conditions and test venues at a more detailed level:

1. 0-Back versus 1-Back
2. 0-Back versus Radio Tuning
3. 0-Back versus Baseline
4. 1-Back versus Radio Tuning

The rationale associated with each of the four selected comparisons is summarized below:

1. 0-Back versus 1-Back. According to Mehler, Reimer, & Dusek (2013), 0-back represents a "very mild" task demand and 1-back represents a "moderate" task demand, which is considered equivalent to a simple conversation. The difference in task demand is considered significant and meaningful and has been the basis of much of the recent work devoted to assessing DRT sensitivity. The ISO working group has selected this difference as one of their main criteria for assessing metric sensitivity. Based on the

results of the ISO joint research initiative and on the omnibus ANOVA results presented above, there is a strong expectation that the DRT performance associated with the 0-back condition will be found to be different from that associated with the 1-Back condition.

2. 0-Back Versus Radio Tuning. Based on our previous work, the demands associated with both 1-back and visual-manual radio tuning were considered acceptable for tasks to be performed while driving. However, the relative positions of 1-back and radio tuning with respect to DRT performance are not well established. This comparison was included to accommodate the possibility that radio tuning might be slightly less disruptive to DRT performance than 1-back task performance, in which case it would be important to determine whether radio tuning was more disruptive to DRT performance than 0-back. Based on the first prediction (0-back versus 1-back) and the assumption that radio tuning will be close to 1-back in its effect on DRT performance (see Prediction 4 below), radio tuning is predicted to be more disruptive to DRT performance than 0-back.
3. 0-Back Versus Baseline. The 0-back condition represents minimally demanding conversation in which the participant needs only to listen and repeat what is being said. While the demands of 0-back are less than those associated with 1-back and radio tuning, they are not trivial. Mehler, Reimer, and Coughlin (2012) showed that the 0-back condition was associated with "measurable changes in heart rate and sweat gland activity, a compensatory reduction in speed, and a constriction of gaze distribution." They concluded that "apparently small increases in demand can add appreciably to the total load on the operator." In a driving protocol, they found that the amount of speed compensation associated with the 0-back condition was identical to that observed in the 1-back condition. Their conclusion is that 0-back represents a real increase in attentional demand that should be detected by a sensitive instrument. Based on these findings and on the omnibus ANOVA results presented above, it is hypothesized that DRT performance will be significantly worse in the 0-back condition than in the baseline (no secondary task) condition.
4. 1-Back Versus Radio Tuning. Both the radio tuning and 1-back tasks are considered as having low to moderate levels of attentional demand; however radio tuning has additional demands associated with the visual-manual task components. The contribution of these additional demands to DRT task performance is not well established. When performed continuously for 1-3 minutes, radio tuning also requires processing of instructions between tasks instances, which may influence DRT performance. The 1-back task allows continuous performance without visual-manual demands or disruption due to transitions between task instances. Thus, while the attentional component of a single instance of radio tuning may generally be comparable to that associated with a 10 to 15-second segment[1] of 1-back task performance, the additional demand associated with the visual-manual components of radio tuning and transitions between task instances is hypothesized to contribute to higher levels of DRT performance degradation for the radio tuning task relative to the 1-back task. However, to the extent that the tactile DRT (TDRT) eliminates the visual interference between DRT and secondary task performance, it is hypothesized that the difference in DRT performance degradation between these tasks will be minimized in the TDRT condition.

[1] Overall mean duration of radio tuning task instances was 10.3 seconds (SD = 3.6) for the non-driving venue and 16.0 seconds (SD = 8.5) for the simulator venue.

Initial analyses were conducted with the full sample (N = 48) and the full (3-minute) data collection interval. The selected set of pairwise comparisons was computed for each of the six combinations of test venue and DRT. The non-parametric Wilcoxon signed-rank test was used for the testing. The test is based on the signs (+ or -) of the differences between the respective mean values; because it does not use the actual metric values, the long tails of the non-normal distributions are therefore effectively attenuated. The Wilcoxon test has power that is essentially equivalent to that of the paired t-test (Myers & Well, 1991). Probability values were adjusted for multiple comparisons to reduce the likelihood of familywise error using the Hochberg adjustment in the MULTTEST procedure in SAS (Westfall, Tobias, & Wolfinger, 2011).

Table 10 presents the adjusted probability values for planned comparisons of mean response time for each combination of test venue and DRT condition using data from all 48 participants and 3-minute data collection intervals.

Table 10. Planned Comparisons: Mean Response Time, N = 48, 3-Minute Data Collection Interval

Comparison		Simulator			Non-Driving		
		HDRT	**TDRT**	**RDRT**	**HDRT**	**TDRT**	**RDRT**
1	0-Back - 1-Back	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001
2	0-Back - Radio Tune	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001
3	0-Back - Baseline	0.0012	<.0001	0.0206	<.0001	<.0001	<.0001
4	1-Back - Radio Tune	<.0001	0.0017	<.0001	<.0001	<.0001	<.0001

For this metric, all differences were statistically significant under all combinations of test venue and DRT. Several test results in the simulator venue are slightly weaker than those in the non-driving venue. In particular, tests with probabilities near p = .05, such as the results of Test 3 in the RDRT simulator condition, are considered weaker than the others. The weakness of this particular result is consistent with Figure 7, which shows the smallest difference in mean response time values between the baseline and 0-back conditions in the RDRT simulator condition. Although they represent statistically significant differences, probability values near p=.05 are more likely to provide inconsistent results when subjected to repeated testing with smaller sample sizes. Table 11 presents the same information for tests performed in the same way using hit rate (proportion of correct responses) as the dependent variable.

Table 11. Planned Comparisons: Mean Hit Rate, N = 48, 3-Minute Data Collection Interval

Comparison		Simulator			Non-Driving		
		HDRT	**TDRT**	**RDRT**	**HDRT**	**TDRT**	**RDRT**
1	0-Back - 1-Back	<.0001	<.0001	<.0001	0.0006	<.0001	0.0011
2	0-Back - Radio Tune	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001
3	0-Back - Baseline	0.0089	0.0188	0.0034	0.0507	0.1497	0.0325
4	1-Back - Radio Tune	<.0001	<.0001	<.0001	0.0695	0.0538	0.7789

Reflecting the differences between the patterns of hit rate means for simulator and non-driving venues apparent in Figure 7, the test results indicate that some of the differences in the non-driving venue were not statistically significant. Specifically, the hit rate differences between 1-back and radio tuning conditions (Comparison 4) were not statistically different for all three non-driving DRT conditions. Similarly, the differences between the 0-back and baseline conditions (Comparison 3) were not statistically significant for two of the three non-driving DRT conditions (the HDRT difference was marginal).

The following two tables present the same statistical test results using data from the shorter 2-minute data collection interval. Results for mean response time are presented in Table 12.

Table 12. Planned Comparisons: Mean Response Time, N = 48, 2-Minute Data Collection Interval

Comparison		Simulator			Non-Driving		
		HDRT	**TDRT**	**RDRT**	**HDRT**	**TDRT**	**RDRT**
1	0-Back - 1-Back	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001
2	0-Back - Radio Tune	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001
3	0-Back - Baseline	0.0086	<.0001	0.0218	<.0001	<.0001	<.0001
4	1-Back - Radio Tune	0.0002	0.0003	<.0001	<.0001	<.0001	<.0001

The outcomes of statistical testing for mean response time reveal that all differences were statistically significant for the 2-minute data collection interval. Moreover, the probability values shown in Table 12 are essentially identical to the values shown in Table 10, which indicates that the use of the 2-minute data collection interval provides identical power and equivalent information relative to the 3-minute data collection interval for this sample size. Results for mean hit rate are presented in Table 13.

Table 13. Planned Comparisons: Mean Hit Rate, N = 48, 2-Minute Data Collection Interval

Comparison		Simulator			Non-Driving		
		HDRT	**TDRT**	**RDRT**	**HDRT**	**TDRT**	**RDRT**
1	0-Back - 1-Back	<.0001	<.0001	<.0001	0.0018	<.0001	0.0020
2	0-Back - Radio Tune	<.0001	<.0001	<.0001	<.0001	<.0001	0.0020
3	0-Back - Baseline	0.0147	0.0873	0.0398	0.0703	0.0156	0.5613
4	1-Back - Radio Tune	<.0001	<.0001	0.0002	0.1211	0.0156	0.6363

The pattern of hit rate test results for the 2-minute data collection interval reveals minor differences from the pattern based on the full 3-minute data collection interval. Here, unlike the results from the longer interval, the 2-minute interval simulator results fail to identify the difference between 0-back and baseline (Comparison 3) for the TDRT simulator condition. In the non-driving test venue, HDRT and RDRT test results indicate no difference for the third and fourth planned comparisons; however, unlike the longer interval results, the non-driving TDRT results reveal significant differences for these tests. The basis for these differences can best be seen in Figure 28.

The following two tables present the Wilcoxon signed-rank test results using data from the shorter 1-minute data collection interval. Results for mean response time are presented in Table 14.

Table 14. Planned Comparisons: Mean Response Time, N = 48, 1-Minute Data Collection Interval

Comparison		Simulator			Non-Driving		
		HDRT	TDRT	RDRT	HDRT	TDRT	RDRT
1	0-Back - 1-Back	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001
2	0-Back - Radio Tune	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001
3	0-Back - Baseline	0.0379	0.0001	0.0068	<.0001	<.0001	<.0001
4	1-Back - Radio Tune	0.0095	0.0005	<.0001	<.0001	<.0001	<.0001

The overall pattern of differences is consistent with those observed for the 2 and 3-minute data collection intervals. Noteworthy is the simulator HDRT comparison number 3, which approaches the boundary of statistical significance (.05) for this time interval. As shown in Figure 11, the magnitude of this difference fluctuates over the 3-minute data collection interval.

Results for mean hit rate are presented in Table 15.

Table 15. Planned Comparisons: Mean Hit Rate, N = 48, 1-Minute Data Collection Interval

Comparison		Simulator			Non-Driving		
		HDRT	TDRT	RDRT	HDRT	TDRT	RDRT
1	0-Back - 1-Back	0.0033	0.0001	<.0001	0.0625	0.0024	0.0337
2	0-Back - Radio Tune	<.0001	<.0001	<.0001	0.0027	<.0001	0.0010
3	0-Back - Baseline	0.0480	0.1038	0.2356	0.0625	0.3594	0.6875
4	1-Back - Radio Tune	<.0001	<.0001	0.0007	0.3456	0.0024	0.0911

Statistical test outcomes, as reflected in probability values, are somewhat weaker in the 1-minute data collection interval hit rate comparisons relative to the 2-minute data collection interval test outcomes. Fewer of the comparisons yielded statistically significant outcomes with data from the 1-minute data collection interval.

The following two tables present the Wilcoxon signed-rank test results using data from the 30-second data collection interval. Mean response time results are presented in Table 16.

Table 16. Planned Comparisons: Mean Response Time, N = 48, 30-Second Data Collection Interval

Comparison		Simulator			Non-Driving		
		HDRT	TDRT	RDRT	HDRT	TDRT	RDRT
1	0-Back - 1-Back	<.0001	0.0019	0.0001	<.0001	<.0001	<.0001
2	0-Back - Radio Tune	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001
3	0-Back - Baseline	0.2816	0.0019	0.1085	<.0001	<.0001	<.0001
4	1-Back - Radio Tune	0.0241	0.0040	<.0001	<.0001	0.0048	<.0001

For Comparison 3, statistical tests found no differences for both the HDRT and RDRT simulator conditions. These results contradict those obtained in all three longer data collection intervals, although it should be noted that some of the earlier differences were relatively weak. Consider the time trends for these conditions shown in Figure 11 and Figure 12. In Figure 11 (HDRT), the effect of shortening the interval from 1 minute to 30 seconds is apparent; the two traces diverge significantly in the second 30 seconds. In contrast, the two traces appear almost parallel in the first two segments in Figure 12 (RDRT), despite the fact that the difference was found to be statistically different in the 1-minute interval, but not significant in the 30-second interval. It may be worth reiterating here that the statistical test is based on the direction of the differences not the magnitude.

Results for mean hit rate are presented in Table 17.

Table 17. Planned Comparisons: Mean Hit Rate, N = 48, 30-Second Data Collection Interval

Comparisons		Simulator			Non-Driving		
		HDRT	TDRT	RDRT	HDRT	TDRT	RDRT
1	0-Back - 1-Back	0.0563	<.0001	0.0038	0.2402	0.1094	0.0004
2	0-Back - Radio Tune	<.0001	<.0001	<.0001	0.0242	0.0015	<.0001
3	0-Back - Baseline	0.1049	0.9214	0.0237	0.2402	0.6250	0.0503
4	1-Back - Radio Tune	0.0010	<.0001	0.0237	0.0651	0.0110	0.0039

The results presented in Table 17 reveal an apparent lack of consistency in the outcomes associated with key comparisons across DRT variants. For Comparison 1 in both venues, the results reveal strong differences in one (non-driving) or two (simulator) of the DRT variants and much weaker differences among the other DRT variants. The strong difference associated with RDRT in the non-driving venue conflicts with non-significant differences found for the other two DRT conditions. In the driving simulator venue, TDRT exhibited a strong difference, while RDRT and HDRT exhibited weaker and marginally non-significant differences, respectively.

The use of 30-second data collection intervals may provide misleading results in some situations. Consider, for example, the time-related traces in Figure 25 (TDRT). Note that the difference in mean hit rates between the baseline and 0-back condition (Comparison 3) in the first 30-second segment is close to zero and that the two traces separate during the second 30-second segment resulting in a much larger difference between these conditions. In Figure 23 (HDRT), the apparently modest difference in the first 30-second segment disappears in the second segment and reverses direction in the third segment. While these changes did not appreciably affect the statistical test outcomes, it is apparent that a 30-second data collection interval may present a snapshot of differences that is not consistent with the experience over longer data collection intervals.

3.7 Effect Size

In previous sections, analyses were focused on four planned comparisons. The objective was to determine which DRT variants in each test venue provided sufficient sensitivity for detecting the targeted differences. These analyses were based on the traditional null hypothesis test significance testing (NHST) model using the traditional criterion of $p < .05$ to define a statistical difference. A high proportion of the statistical results revealed differences using this criterion, which indicated that all test conditions were successful in detecting the targeted differences; however, this outcome provided little useful information to distinguish among the various test conditions. The present analyses examined the effect sizes (ESs) associated with the planned comparisons in each of the testing situations on the assumption that larger ES values reflect greater sensitivity for a given test situation in which the population ES is assumed to be constant.

Although it is possible to construct a non-parametric ES which would provide better continuity with the Wilcoxon signed-rank test, this undertaking was beyond the scope of the current effort. Therefore, Cohen's d was computed to represent ES for the targeted pairwise differences for each combination of test venue and DRT. Cohen's d is a standardized difference between group means, defined in standard deviation (SD) units (Kelley & Preacher, 2012; Cumming, 2014). There is some disagreement concerning the proper term to use in the denominator when the data come from a repeated-measures design. In particular, Cohen (1988) recommended using the SD_{diff} for this purpose, while more recently Cumming (2013) recommended using the pooled standard deviation (SD_{av}) as would normally be done when the design involved two independent groups. For the current purposes, SD_{diff} was selected as the standardizer because the use of this term is consistent with the structure of the paired-t test, which is appropriately used to analyze data from repeated-measures designs, and is a term that reflects the experimental goal of assessing the sensitivity of DRT variants.

The lack of normality in the original data can create problems for estimating the accuracy of Cohen's d. Adjustments to confidence intervals (CIs) have been proposed as a means to compensate for deviations from normality in the underlying data (Kelley, 2005). This problem was addressed in the present work by the computation of bootstrap CIs, which involved creating a distribution of Cohen's d values using repeated random samples from the observed data. For this effort, bias corrected and accelerated (BCa) CIs were used (Canty & Ripley, 2014; Davison & Hinkley, 1997; R Core Team, 2014). BCa CIs are generally preferred over percentile CIs (Kelley, 2005). The bootstrap procedure makes no assumptions about the distributions of the parent population or sample except that the data were randomly sampled and representative of the parent population. The width of the CI reflects the precision of the ES estimate.

The following figures (Figures 40-42) present the ES values and their associated CIs for three data collection intervals (3, 2, and 1 minute) and two test venues (driving simulator, non-driving). Each figure consists of four blocks (left to right), one for each of the planned comparisons defined in Section 3.6. For each planned comparison, ES values are presented for each of the three DRT variants. It should be reiterated that the data collection intervals are not independent; data from the first minute are included in each interval; data from the second minute are included in the 2- and 3-minute intervals.

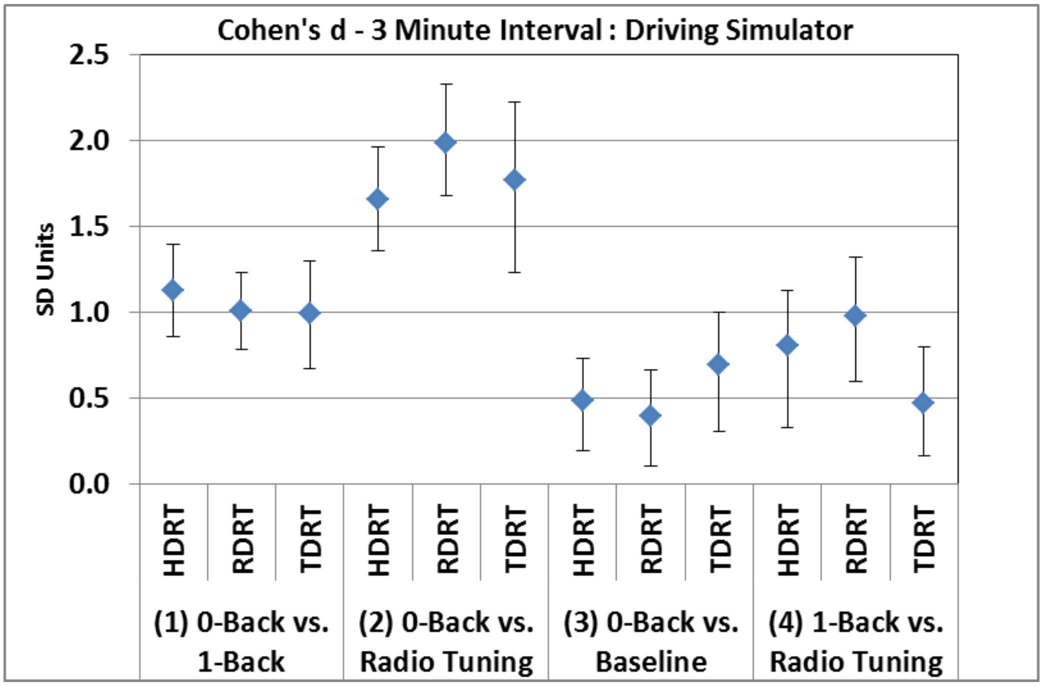


Figure 40. Effect Sizes and Confidence Intervals by Secondary Task and DRT: 3-Minute Data Collection Interval in the Driving Simulator

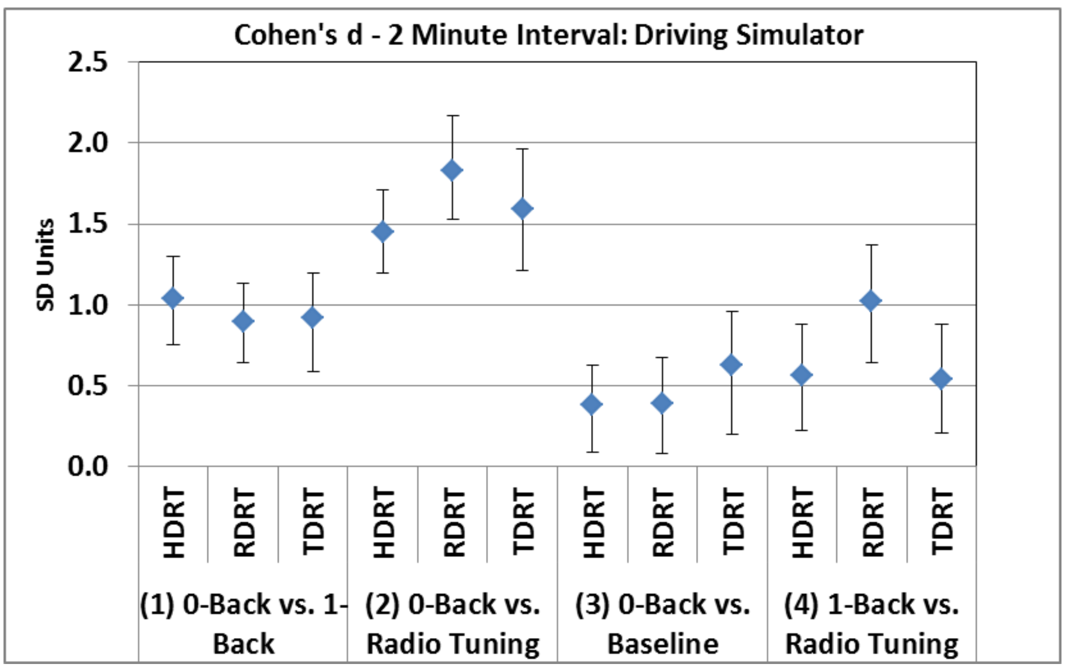


Figure 41. Effect Sizes and Confidence Intervals by Secondary Task and DRT: 2-Minute Data Collection Interval in the Driving Simulator

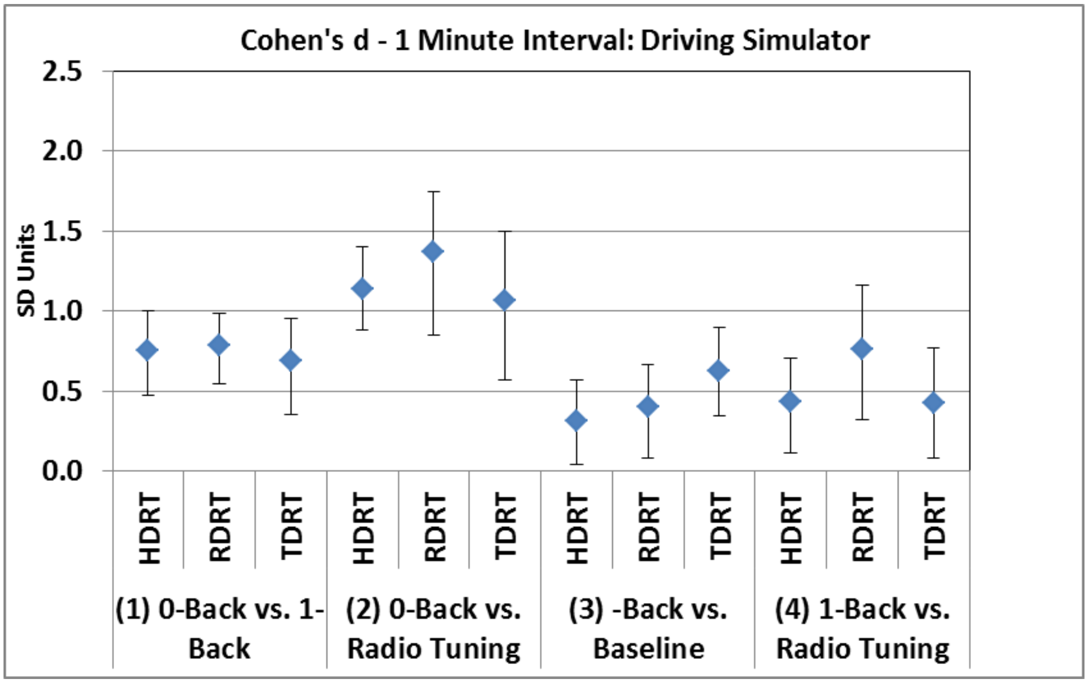


Figure 42. Effect Sizes and Confidence Intervals by Secondary Task and DRT: 1-Minute Data Collection Interval in the Driving Simulator

For each data collection interval, Comparison 2 (0-Back versus Radio tuning) had the largest ES values, while Comparison 3 (0-Back versus Baseline) had the smallest ES values. The figures reveal some differences among the DRT variants. In particular, the ES values for RDRT appear elevated relative to the other two DRT variants for Comparisons 2 and 4.

Radio tuning was the only secondary task condition in the experiment that had visual and manual task demands, which required participants to look away from the RDRT target location, thus providing an opportunity for RTs to be influenced not only by attentional task demands but also by delay due to the RDRT target not always being visible when it first appeared. HDRT and TDRT variants incorporated controls to ensure that targets were always perceptible when first activated. Thus for Comparisons 2 and 4, the elevated RDRT ES values are consistent with the interpretation that RDRT RTs in the radio tuning condition include contributions from both attentional task demands and participants' failure to detect the target immediately upon activation, while n-back condition RTs include only the attentional component. According to this interpretation, the ES values for HDRT and TDRT for these same comparisons are smaller because the RTs for one condition do not include the additional time required to initially detect the target. If this interpretation is correct, the RDRT ES values are not directly comparable to those obtained using HDRT and TRDT methods for comparisons between auditory-vocal and visual-manual tasks.

For Comparisons 1 and 3, RDRT ES values were more consistent with those associated with the other DRT variants. For Comparison 3, which represents the most challenging difference, the TDRT had slightly elevated ES values across all three data collection intervals. For Comparison 1, HDRT had slightly elevated ES values relative to the other two DRT variants for the longer two data collection intervals. ES values for each of the four planned comparisons are presented by DRT variant for the non-driving test venue in Figure 43 to Figure 45.

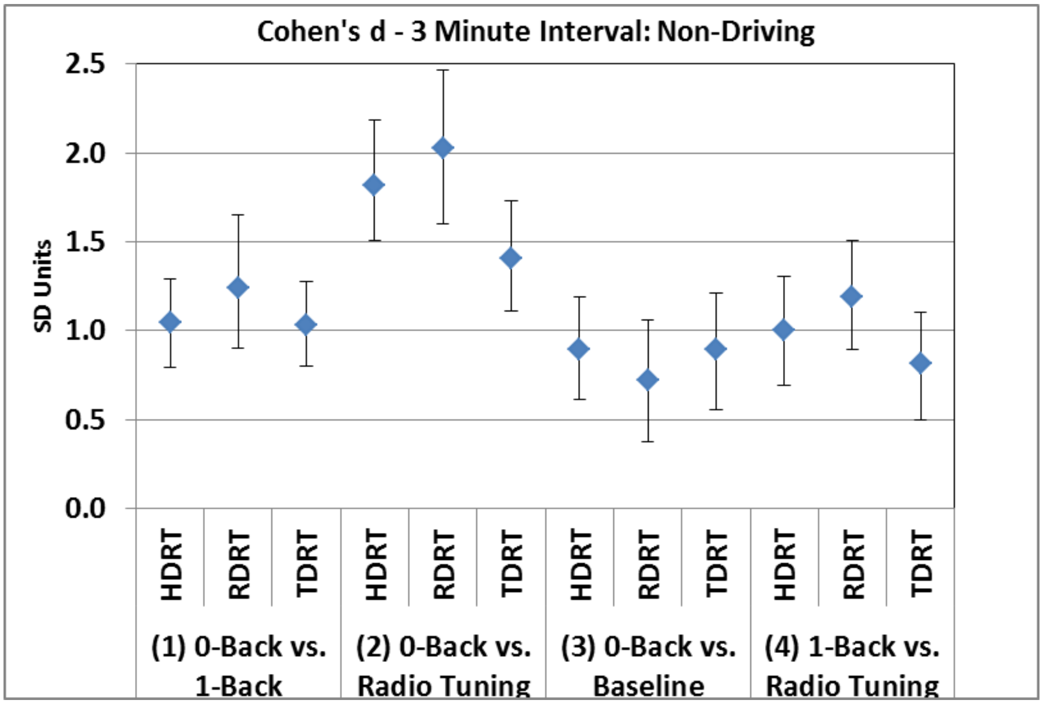


Figure 43. Effect Sizes and Confidence Intervals by Secondary Task and DRT: 3-Minute Data Collection Interval in the Non-Driving Venue

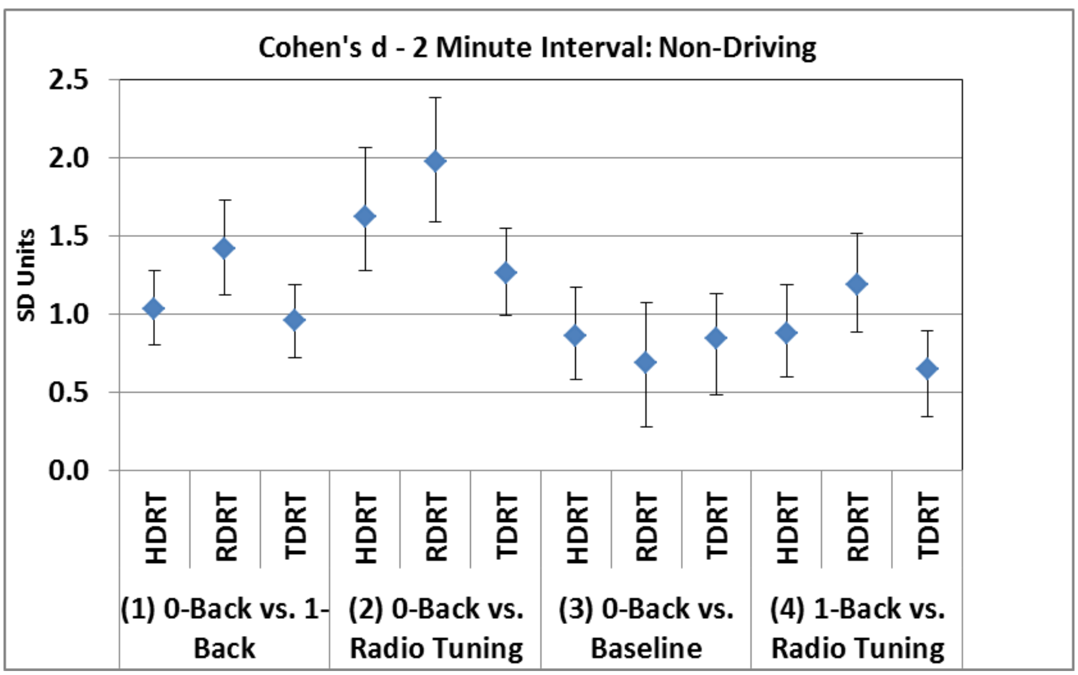


Figure 44. Effect Sizes and Confidence Intervals by Secondary Task and DRT: 2-Minute Data Collection Interval in the Non-Driving Venue

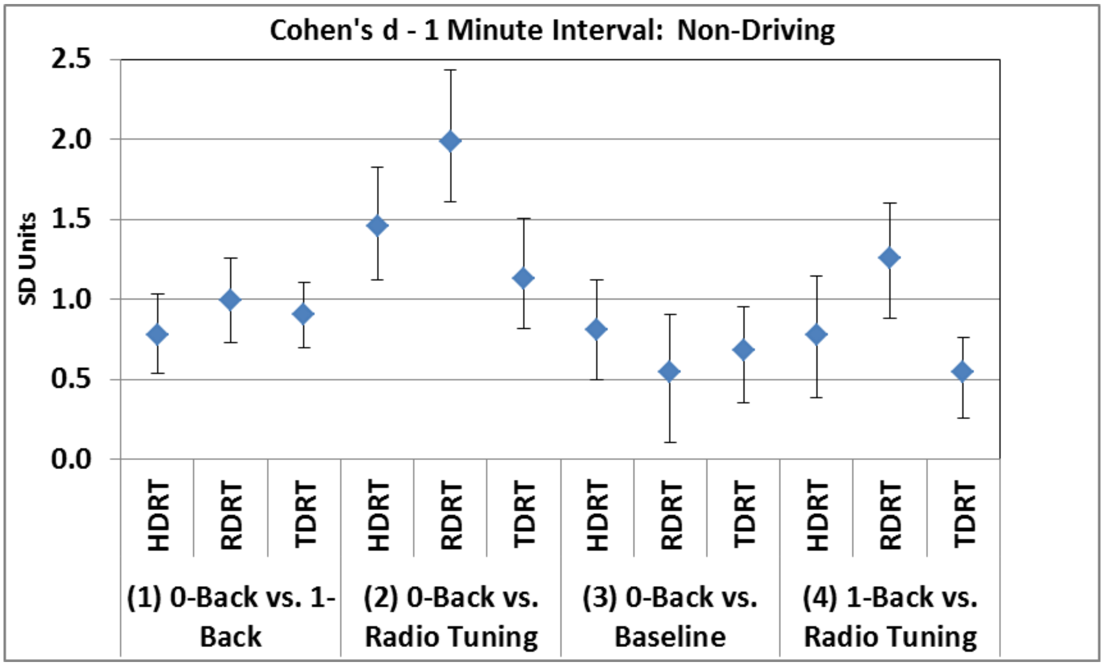


Figure 45. Effect Sizes and Confidence Intervals by Secondary Task and DRT: 1-Minute Data Collection Interval in the Non-Driving Venue

Overall, ES values appear slightly larger in the non-driving venue then in the driving simulator venue, reflecting increased sensitivity due to reduced variability in the absence of a concurrent driving task. The pattern of elevated ES values for Comparisons 2 and 4 for RDRT observed in the driving simulator data was also apparent for the non-driving data, supporting the same interpretation as described above for the driving simulator data. However, unlike the simulator data, this pattern was also apparent for Comparison 1 in the non-driving data, which cannot readily be explained by the potential confounds described above. The non-driving ES values for TDRT were noticeably smaller than the HDRT values for Comparisons 2 and 4. In addition, the higher ES values for TDRT Comparison 3 observed in the driving simulator data were not apparent in the non-driving data. Together, these observations suggest less relative sensitivity for the TDRT in the non-driving venue.

While it would generally be appropriate to associate elevated ES values as evidence of increased test sensitivity, the potential confound discussed above argues against this interpretation for Comparisons 2 and 4 with the RDRT. Among the four planned comparisons, Comparisons 1 and 3 were established as the main criteria for assessing the DRT variants. For these comparisons, differences in ES values among the DRT variants were relatively small. In the simulator data, the ES values associated with HDRT were slightly larger than those for the other DRT variants for Comparison 1. TDRT had slightly larger ES values than other DRT variants for Comparison 3. In the non-driving test venue, the RDRT had slightly elevated ES values for Comparison 1, while the HDRT and TDRT had slightly elevated ES values for Comparison 3.

Most generally, the ES values were larger at the longer data collection intervals. To better demonstrate these trends, the following figures present the same data organized so that data from all data collection intervals are presented together. To accommodate this presentation, separate figures are presented for each of the four planned comparisons in each test venue.

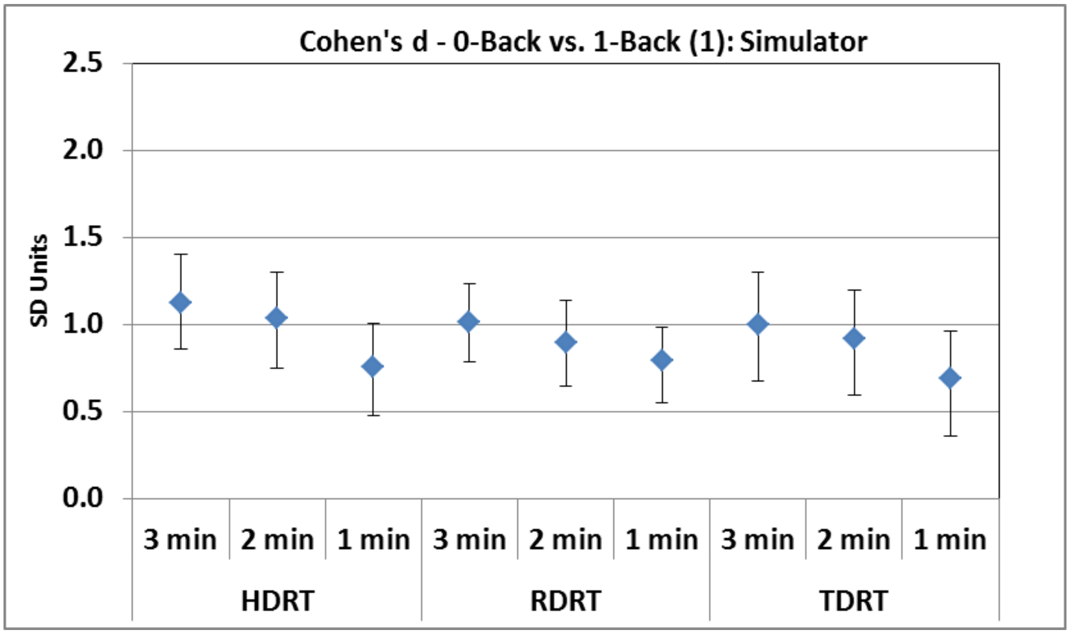


Figure 46. Effect Sizes and Confidence Intervals by Data Collection Interval and DRT: Driving Simulator Comparison 1

For Comparison 1 in the driving simulator, all three DRT variants reveal progressive declines in ES with shorter data collection intervals. As mentioned above, the ES values were slightly larger for the HDRT than for the other DRT variants.

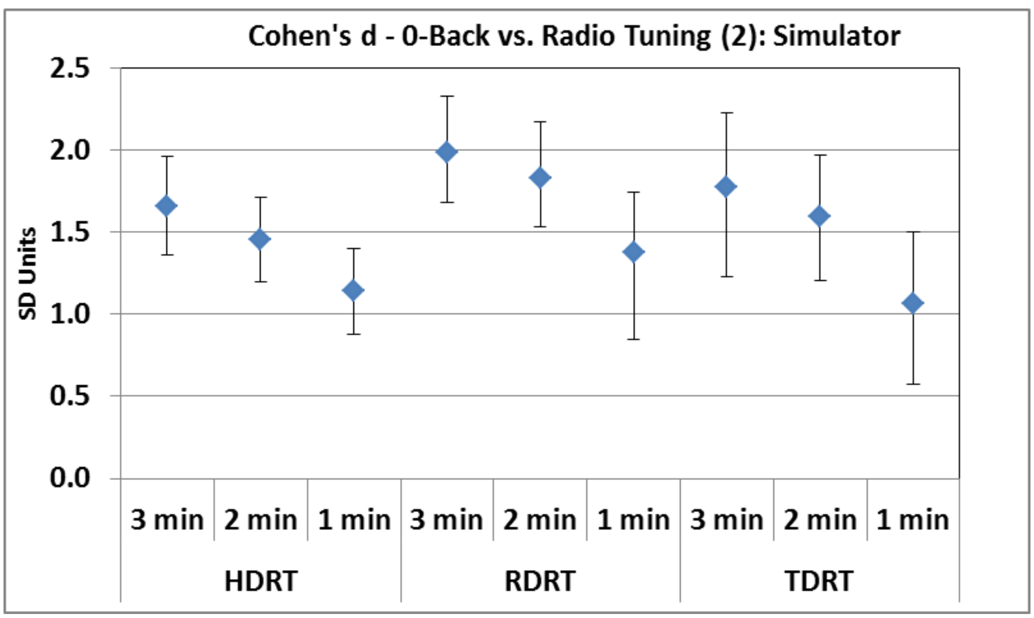


Figure 47. Effect Sizes and Confidence Intervals by Data Collection Interval and DRT: Driving Simulator Comparison 2

For Comparison 2 in the driving simulator, all three DRT variants reveal progressive declines in ES with shorter data collection intervals. The declines were somewhat sharper between the 2- and 1-minute time intervals for the RDRT and TDRT conditions. The RDRT ES values were greater than those of other DRT variants at all three time intervals; however, this may have been due to the above-mentioned confound. Otherwise, ES values for TDRT were slightly greater than those for HDRT at the two longer data collection intervals; however the considerably larger CI width indicates increased variability in the TDRT data.

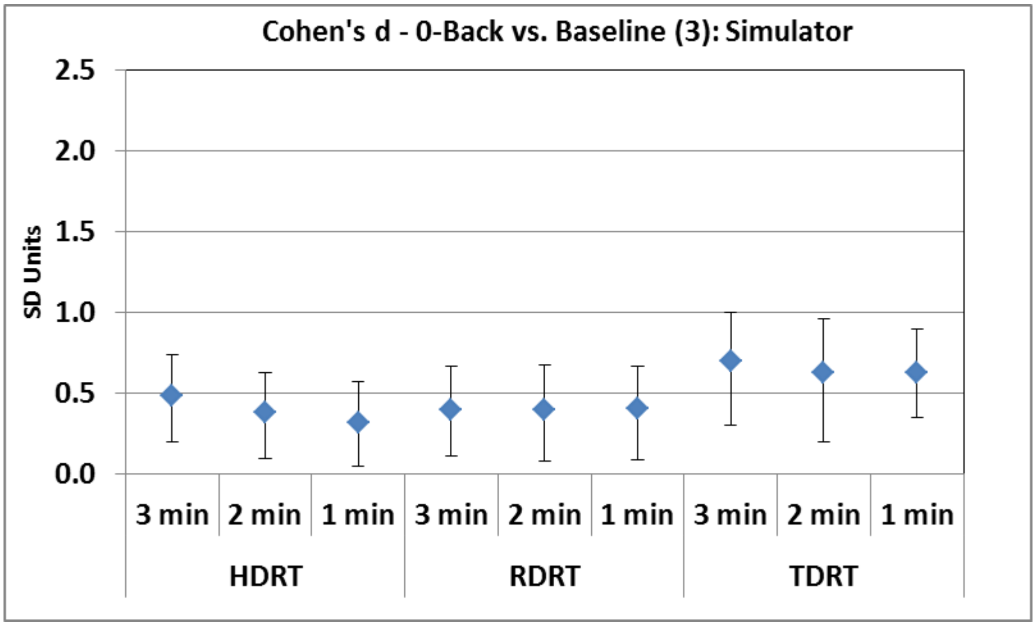


Figure 48. Effect Sizes and Confidence Intervals by Data Collection Interval and DRT: Driving Simulator Comparison 3

The relatively small ES values associated with Comparison 3 in the driving simulator were generally consistent across time intervals for RDRT. The HDRT ES values reveal a progressive albeit small decrease as the data collection interval becomes shorter. TDRT was associated with slightly larger ES values and slightly wider CIs for this comparison.

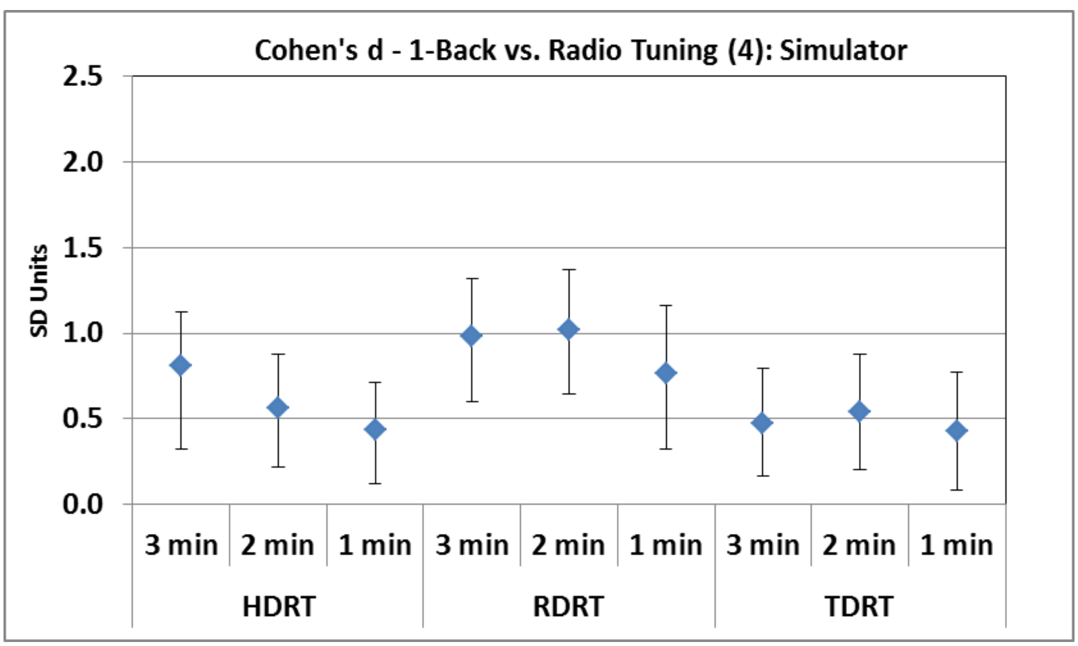


Figure 49. Effect Sizes and Confidence Intervals by Data Collection Interval and DRT: Driving Simulator Comparison 4

For Comparison 4, the relatively large ES values for RDRT should be interpreted with caution due to the above-mentioned confound. The HDRT ES values were similar to those for TDRT at the two shorter intervals, but larger at the longest interval. Unlike the other two DRT variants, the HDRT ES values exhibited a progressive decline with decreasing interval duration.

The following figures present ES data for the non-driving venue organized for viewing trends associated with data collection interval duration.

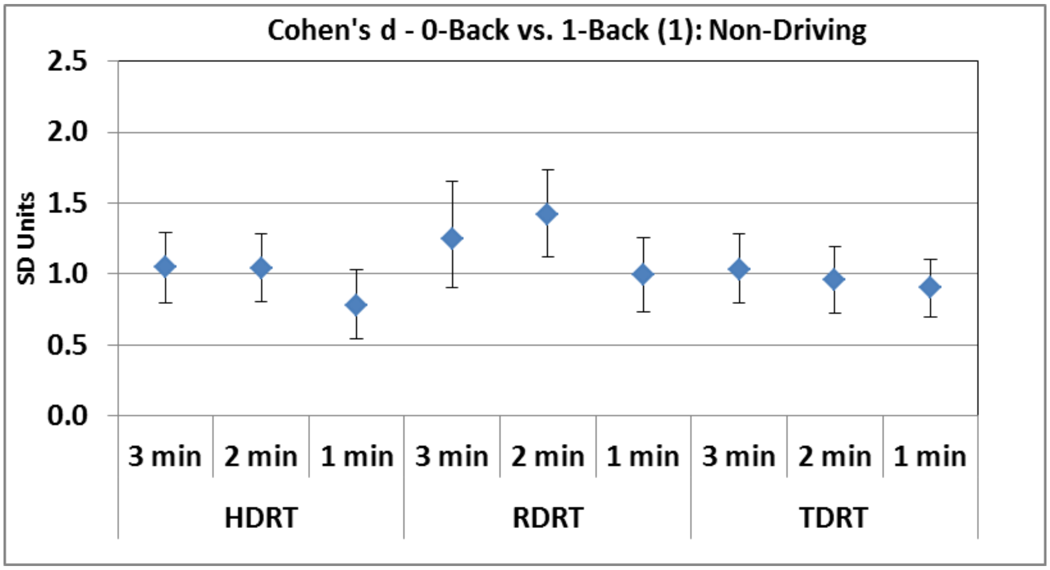


Figure 50. Effect Sizes and Confidence Intervals by Data Collection Interval and DRT: Non-Driving Comparison 1

For Comparison 1 in the non-driving venue, RDRT had slightly larger ES values along with wider CIs relative to the other two DRT variants. HDRT and RDRT ES values dropped off at the shortest time interval; TDRT ES values exhibited a slight progressive decline in ES value across the 3 time intervals.

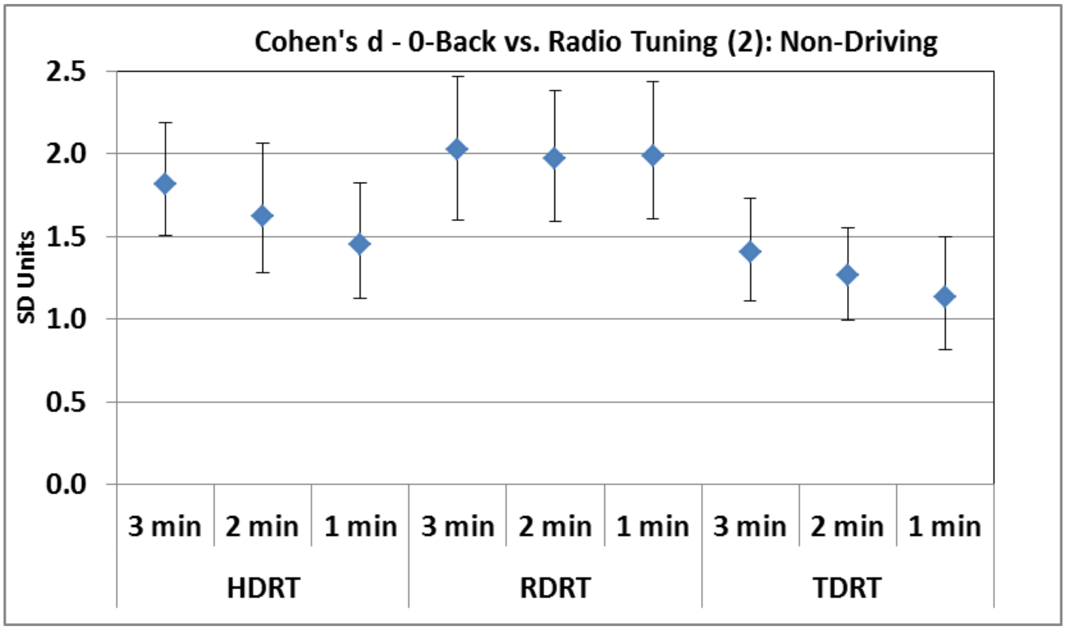


Figure 51. Effect Sizes and Confidence Intervals by Data Collection Interval and DRT: Non-Driving Comparison 2

For Comparison 2, both HDRT and TDRT exhibited progressive declines in ES values with decreasing time interval. HDRT ES values were consistently greater than TDRT ES values. The elevated RDRT ES values were consistent across the data collection intervals and had slightly wider CI values, particularly relative to TDRT.

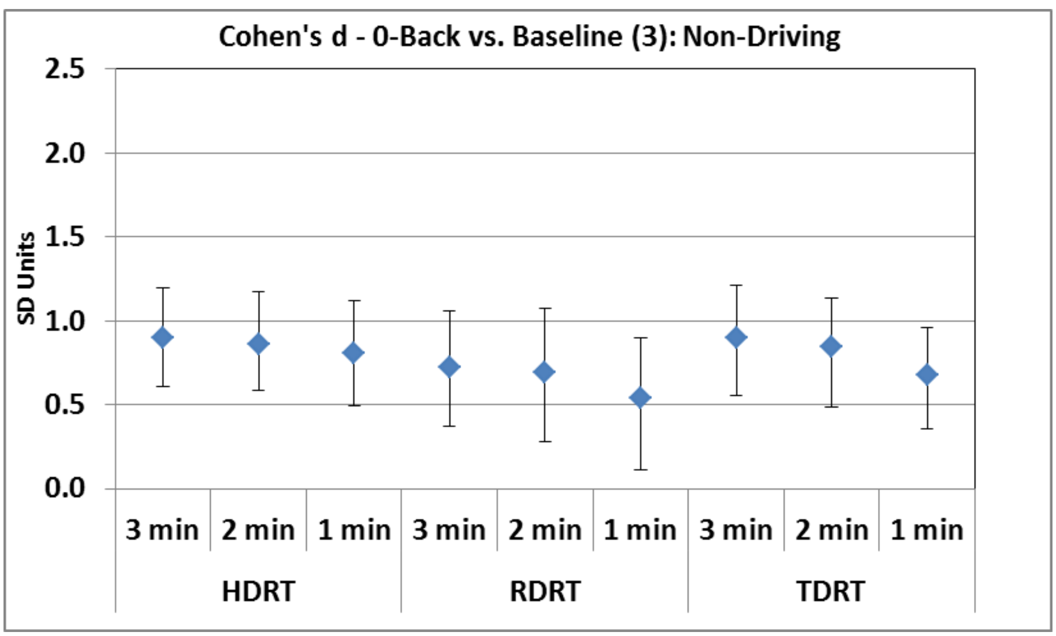


Figure 52. Effect Sizes and Confidence Intervals by Data Collection Interval and DRT: Non-Driving Comparison 3

For Comparison 3, ES values decreased progressively for all three DRT variants with decreasing time intervals. This effect was most prominent between the 1- and 2- minute intervals. HDRT and TDRT ES values were essentially equivalent at the longer two data collection intervals. RDRT ES values were smaller than those for HDRT and TDRT. For this planned comparison, ES values from the non-driving venue were consistently greater than those associated with the driving simulator. Magnitudes of differences were largest for this comparison. Simulator ES values were consistently smaller than 0.5 for HDRT and TDRT; the corresponding non-driving venue ES values, shown in this figure, were consistently between 0.5 and 1.0. TDRT ES values did not reveal similar differences between test venues for this comparison.

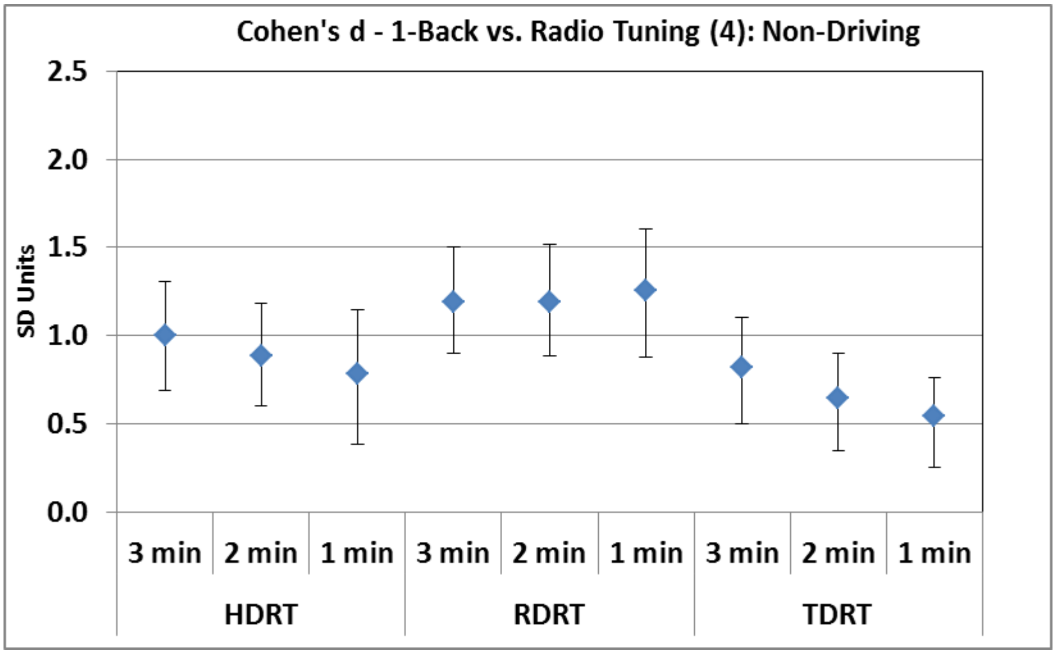


Figure 53. Effect Sizes and Confidence Intervals by Data Collection Interval and DRT: Non-Driving Comparison 4

The progressive time-related deterioration in ES values was apparent for both HDRT and TDRT, with HDRT values being modestly greater than TDRT across all time intervals. RDRT ES values did not decrease at shorter data collection intervals.

3.8 Small Sample Statistical Test Results

The experiment was designed to provide comparable data from two independent samples of 24 participants, each of which was selected according to the Phase 1 NHTSA Guidelines specifications. These two samples are referred to as the small samples. The experiment was administered in the same way to each of the two groups, with one exception, namely the order of test venues. One group had the simulator test venue first while the second group had the non-driving test venue first. Preliminary examination of the data did not reveal any time-related trends in DRT performance that would be consistent with progressive deterioration of performance due to cumulative fatigue over the course of the test day. We therefore concluded that the samples were comparable. Tables 18 and 19 provide summaries of the statistical test outcomes for each small sample in each combination of DRT and test venue, using data collection intervals of 1 and 2 minutes. Statistical tests were performed using the Wilcoxon signed-rank test for the family of planned comparisons defined above. Within each combination of test venue and data collection interval, Sample columns labeled Test and Retest present the probability values associated with the statistical tests for the two small samples. Table 18 presents results for mean response time and Table 19 presents results for mean hit rate.

Table 18. Statistical Test Results for Small Sample Tests: Mean Response Time

Comparisons		Simulator				Non-Driving			
		2 minute		1 minute		2 minute		1 minute	
		Sample (N = 24)							
	HDRT Mean RT	**Test**	**Retest**	**Test**	**Retest**	**Test**	**Retest**	**Test**	**Retest**
1	0-Back - 1-Back	<.0001	<.0001	0.0026	0.0002	<.0001	<.0001	<.0001	0.0001
2	0-Back - Radio	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001
3	0-Back - Baseline	0.0428	0.0921	0.1487	0.1327	<.0001	<.0001	<.0001	0.0002
4	1-Back - Radio	0.0050	0.0447	0.0530	0.1327	0.0001	<.0001	<.0001	0.0002
	TDRT Mean RT	**1**	**2**	**1**	**2**	**1**	**2**	**1**	**2**
1	0-Back - 1-Back	0.0014	<.0001	0.0497	0.0001	<.0001	<.0001	<.0001	<.0001
2	0-Back - Radio	<.0001	<.0001	0.0030	<.0001	<.0001	<.0001	<.0001	<.0001
3	0-Back - Baseline	0.0171	<.0001	0.0497	0.0015	<.0001	0.0026	0.0014	0.0258
4	1-Back - Radio	0.0114	0.0171	0.0497	0.0051	<.0001	0.0497	<.0001	0.1849
	RDRT Mean RT	**1**	**2**	**1**	**2**	**1**	**2**	**1**	**2**
1	0-Back - 1-Back	0.0007	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	0.0004
2	0-Back - Radio	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001
3	0-Back - Baseline	0.0534	0.1973	0.0534	0.0717	<.0001	0.0051	0.0009	0.0096
4	1-Back - Radio	<.0001	<.0001	<.0001	0.0213	<.0001	<.0001	<.0001	<.0001

In the driving simulator venue, Test 3 (0-back versus Baseline) was associated with probability values that did not meet the statistical test criterion ($p < .05$). This pattern is consistent with the consistently smaller ES values associated with this test. In the driving simulator venue, the TDRT was more sensitive to this difference than the other two DRT variants. Statistically significant differences were observed for the TDRT for both samples at both of the data collection intervals; these differences were consistent across the two samples, however one of the 1-minute differences was relatively weak. In the non-driving test venue, the statistical test outcomes were consistent and unequivocal; differences associated with all four selected comparisons were statistically significant, and with the exception of several relatively weak effects associated with Comparisons 3 and 4 in the TDRT condition, the effects were consistently strong.

Test results presented in Table 19 also provide information about the test-retest reliability of the test outcomes with smaller sample sizes. Test-retest reliability is defined as the agreement of statistical outcomes between the two small samples for a particular test. Generally, the test results were consistent across the two samples indicating high test-retest reliability. Exceptions, which are highlighted, included the difference in test outcomes observed in the 2-minute data collection interval for Comparison 3 in HDRT (simulator) and the difference in the 1-minute data collection interval for Comparison 4 in TDRT (non-driving). This latter difference is noteworthy because neither probability value is close to the criterion value, which generally indicates that there is less uncertainty about the interpretation of the test outcomes.

Table 19. Statistical Test Results for Small Sample Tests: Mean Hit Rate

Comparisons		**Simulator**				**Non-Driving**			
		2 minute		**1 minute**		**2 minute**		**1 minute**	
		Sample (N = 24)							
	HDRT Mean Hit Rate	**Test**	**Retest**	**Test**	**Retest**	**Test**	**Retest**	**Test**	**Retest**
1	0-Back - 1-Back	0.0042	0.0002	0.0344	0.0524	0.0469	0.1324	0.1816	0.5000
2	0-Back - Radio	<.0001	<.0001	0.0008	<.0001	0.0021	0.0173	0.1172	0.1094
3	0-Back - Baseline	0.4579	0.0112	0.5723	0.0020	0.5000	0.3438	0.3372	0.5000
4	1-Back - Radio	0.0027	0.0006	0.0255	<.0001	0.4160	0.4421	0.3372	0.7689
	TDRT Mean Hit Rate	**1**	**2**	**1**	**2**	**1**	**2**	**1**	**2**
1	0-Back - 1-Back	0.0005	<.0001	0.0005	0.0935	0.0117	0.0022	0.2500	0.0352
2	0-Back - Radio	<.0001	<.0001	<.0001	<.0001	0.0005	0.0022	0.0015	0.0322
3	0-Back - Baseline	0.3630	0.1199	0.1762	0.2687	1	0.0625	1.0000	0.3750
4	1-Back - Radio	<.0001	0.0015	<.0001	0.0002	0.0212	0.3289	0.0015	0.4534
	RDRT Mean Hit Rate	**1**	**2**	**1**	**2**	**1**	**2**	**1**	**2**
1	0-Back - 1-Back	0.0510	0.0005	0.0234	0.0027	0.2139	0.0109	0.2813	0.3750
2	0-Back - Radio	<.0001	<.0001	<.0001	0.0006	0.2139	0.0315	0.0625	0.0938
3	0-Back - Baseline	0.3397	0.0552	0.7930	0.0396	0.5625	0.9531	1.0000	0.5000
4	1-Back - Radio	0.0047	0.0252	0.0029	0.0396	0.5625	0.9920	0.3750	0.3750

Test results presented in Table 19 for mean hit rate reveal some equivocation, with different test outcomes for different samples in the same test conditions. Highlighted pairs indicate test outcomes based on the $p < .05$ statistical criterion for which the two samples did not agree; in particular, the statistical test outcome was statistically significant for one sample and not significant for the other sample. The lack of agreement, indicative or lower test-retest reliability is more apparent for hit rate than for response time (Table 18).

In an attempt to consolidate the information from the previous two tables, the following tables (Tables 20 and 21) summarize the information provided by the statistical tests performed on small (N = 24) samples using data from 1- and 2-minute data collection intervals. Following the rationale provided at the outset, the two diagnostic comparisons (Comparisons 1 (0-back versus 1-back) and 3 (0-back versus baseline)) are used to determine sensitivity. As has been evident throughout, Comparison 1 was relatively easy to detect while Comparison 3 posed a significant challenge and thus provided a stronger test of the sensitivity of the metrics. This presentation allows test outcome data from both response time and hit rate to be combined. Data are provided separately for the two test venues. Following is a summary of the information presented:

1. Statistical test outcomes – Frequencies outside parentheses in each table cell report the number of statistically significant outcomes for each of the 4 selected comparisons. The four comparisons include the two selected comparisons for the response time metric (RT-1 and RT-3) and the two selected comparisons for the hit rate metric (HR-1 and HR-3).

The maximum possible score for each cell is two, which would indicate that the testing for both samples detected a significant difference. The maximum Total score is 8.

2. The Agreement row summarizes the consistency (i.e., test-retest reliability) of the test outcomes for the two samples. The denominator indicates the number of tests that found the difference to be statistically significant in either of the samples. The numerator indicates the number of tests that had consistent results among those that detected a difference in at least one of the samples.

Table 20. Summary of Statistical Test Outcomes for Small Sample Testing on Driving Simulator Test Venue

	HDRT		TDRT		RDRT	
Comparison	**2 min**	**1 min**	**2 min**	**1 min**	**2 min**	**1 min**
Simulator RT-1	2	2	2	2	2	2
Simulator RT-3	1	0	2	2	0	0
Simulator HR-1	2	1	2	1	1	2
Simulator HR-3	1	1	0	0	0	1
Total	6	4	6	5	3	5
Agreement	2/4	1/3	3/3	2/3	1/2	2/3

Considering the total number of differences detected (Total row in Table 20), both the HDRT 2-minute test condition and the TDRT 2-minute test condition detected six of eight possible differences. The level of agreement among the test outcomes for the two samples was better in the TDRT 2-minute condition (3/3) than in the HDRT 2-minute condition (2/4). Based on these small-sample criteria, the TDRT 2-minute test protocol represents the best performing among those in the driving simulator venue. One minor caveat is that HDRT did reveal slightly greater sensitivity to the HR-3 difference than TDRT; however, this difference was weak and not detected consistently in both samples at either data collection interval.

Table 21. Summary of Statistical Test Outcomes for Small Sample Testing on Non-Driving Test Venue

	HDRT		TDRT		RDRT	
Comparison	**2 min**	**1 min**	**2 min**	**1 min**	**2 min**	**1 min**
Non-Driving RT-1	2	2	2	2	2	2
Non-Driving RT-3	2	2	2	2	2	2
Non-Driving HR-1	1	0	1	0	1	0
Non-Driving HR-3	0	0	0	0	0	0
Total	5	4	5	4	5	4
Agreement	2/3	2/2	2/3	2/2	2/3	2/2

Looking first at the total number of differences detected in the non-driving venue small sample tests, three of the six conditions, including all three of the 2-minute data collection interval tests, had five of eight differences detected. The corresponding 1-minute data collection interval tests all had four of eight differences detected, reflecting a slight loss of sensitivity at the shorter interval. The agreement scores were identical for these three conditions (2/3). Agreement ratios for the 1-minute intervals all had smaller denominators than the corresponding 2-minute interval values, due to the decreased sensitivity and inability to detect more differences with the shorter data collection intervals. Among the more sensitive 2-minute conditions, the only difference observed among the DRT variants was the slightly stronger difference observed in the RT-3 test for HDRT (4+4=8), relative to the other two conditions (TDRT: 4+2=6; RDRT: 4+2=6). Thus for the non-driving test venue, the 2-minute data collection interval revealed a consistent if modest increase in sensitivity relative to the 1-minute interval. Among DRT variants, the HDRT has a slight advantage over the other two 2-minute test results based on the small sample statistical test results.

3.9 Secondary Analyses

Newer DRT variants, including HDRT and TDRT, have incorporated modifications to control for the potential conflict inherent in the RDRT, which makes it difficult to determine the relative contributions of attentional versus visual load to DRT performance degradation for tasks involving visual-manual demands. HDRT addresses this problem by keeping the visual target location fixed relative to the driver's head while TDRT eliminates the visual stimulus altogether. Both approaches thus control for differences in drivers' head positions at target onset that influence remote target visibility and thus DRT performance. Accordingly, if RDRT performance is affected by both visual and attentional task demands, and other DRT variants are influenced only by attentional task demands, it would follow that the difference in DRT performance between a task with no visual demands (e.g., 1-back) and one that has visual demands (e.g., Radio Tuning) would be greater for RDRT than for the other two DRT conditions. The following figures present mean response times and hit rates for these two secondary task conditions in each DRT condition for both test venues.

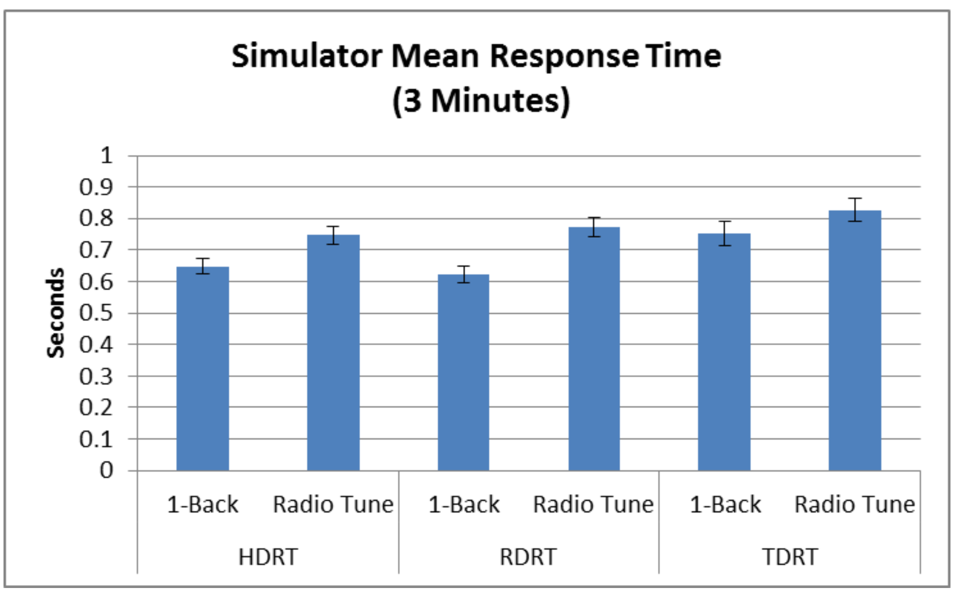


Figure 54. Simulator Mean Response Times: 3-Minute Data Collection Interval

The difference between the two task conditions shown in Figure 54 is largest in the RDRT condition, which is consistent with the hypothesis that the increase in performance degradation in the RDRT radio tuning task condition relative to the 0-back condition includes time due to the increased attentional demands of the radio tuning task and that the target is not immediately visible when it appears while the driver is looking at the radio. According to this proposed explanation, the smaller differences between these two tasks observed in the HDRT and TDRT conditions reflect the fact that this latter component does not contribute to response times in the HDRT and TDRT conditions. As shown in Figure 55, this pattern is not immediately evident in the response time means obtained in the non-driving test venue. In this venue, the difference between 1-back and radio tuning response time means is slightly greater in the RDRT condition than in the other two conditions, but the difference is very small. If the hypothesized phenomenon does exist, these results suggest that it is present only in the driving simulator test venue, which differs from the non-driving venue based on the multi-tasking requirement, which is not part of the non-driving test protocol.

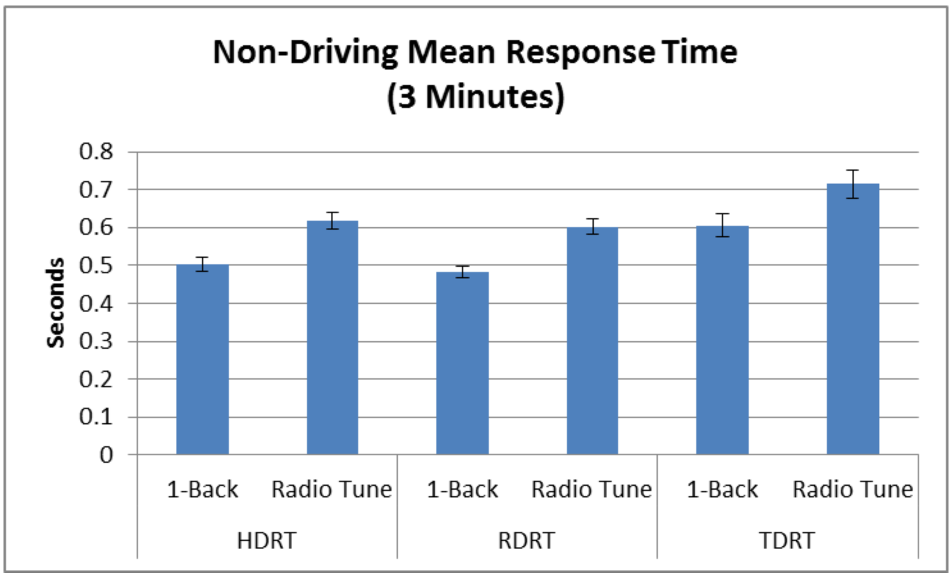


Figure 55. Non-Driving Mean Response Times: 3-Minute Data Collection Interval

This phenomenon could also affect the hit rate in the RDRT condition, such that the difference in hit rates between these two secondary task conditions would be greater in the RDRT condition than in the other two DRT conditions. Figure 56 presents the mean hit rate values from the 3-minute data collection intervals for the driving simulator venue in which participants performed the 1-back and the radio tuning trials.

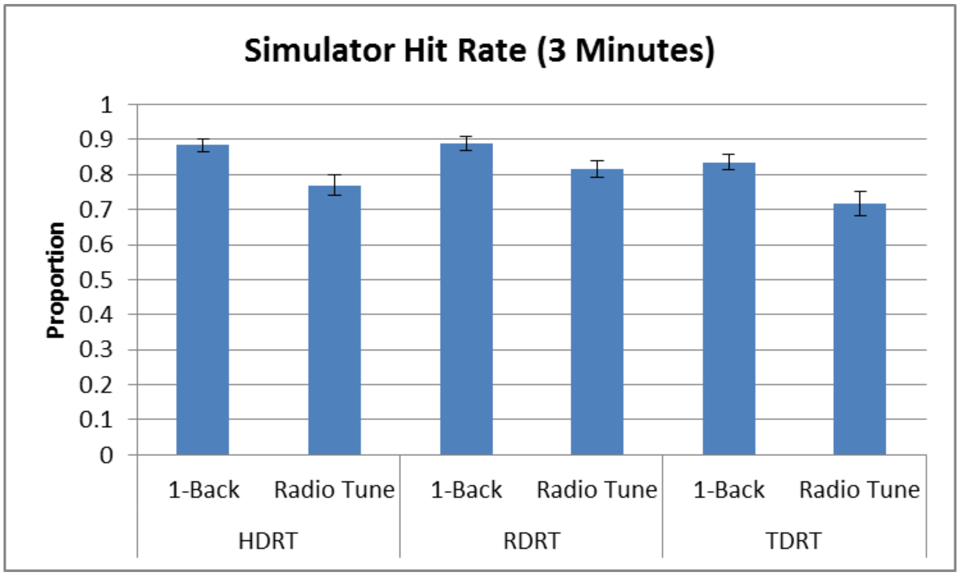


Figure 56. Simulator Mean Hit Rates: 3-Minute Data Collection Interval

The mean hit rate values summarized over the 3-minute data collection interval are not consistent with the hypothesis presented above. Radio tuning hit rate did not suffer in the RDRT condition relative to the other DRT conditions and the difference in hit rate between the two secondary task conditions is not greater in the RDRT condition than in the other two DRT conditions. It appears that the potential conflict inherent in the RDRT condition did not contribute to hit rate. Rather, the pattern of results, including the higher mean hit rates and the slightly smaller difference between secondary task conditions suggests that the RDRT signals were more readily detectable than those associated with the other two DRT variants.

Figure 57 presents the same means for the non-driving test conditions. Differences between conditions in mean hit rate values were smaller due to the aforementioned ceiling effects, however noteworthy in this presentation is the observation that the difference in mean hit rate values between these two conditions was essentially zero for the RDRT condition, which suggests that the remote target was generally much easier to detect that the other two targets in the radio tuning task condition. This pattern was generally consistent across the two test venues.

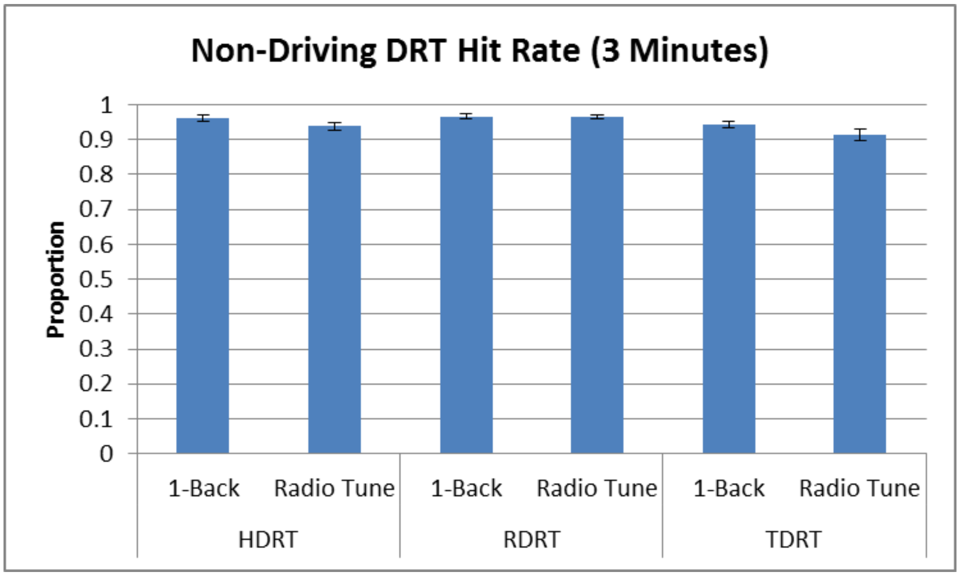


Figure 57. Non-Driving Mean Hit Rates: 3-Minute Data Collection Interval

A second question concerns the effect of differences in secondary task pacing. The n-back conditions were externally-paced, which controls the demands of the task. In contrast, the radio tuning task was self-paced, which allows the participants freedom in determining the pace and thus the level of task demands. Faster self-paced secondary task performance, reflected as more task instance completions during a fixed data collection interval, implies a (self-imposed) higher level of attentional demand than slower performance. If increased secondary task performance is found to be associated with decreased DRT performance, this may be evidence that secondary task load is not consistent across conditions or participants, which may raise questions about the validity of comparisons between DRT effects on self-paced tasks versus externally-paced tasks. It is therefore of interest to determine whether differences in the number of radio tuning task instances completed during the 3-minute data collection interval affected DRT performance.

The number of completed radio tuning task instances varied systematically by test venue. On average, participants completed approximately nine task instances over the 3-minute data collection interval in the non-driving test venue versus approximately 6.5 task instances over the same duration in the simulator test venue. There were no differences among the three DRT conditions in each venue. To examine the possibility that the self-selected pace of secondary task performance on radio tuning trials affected DRT performance, correlations among DRT response time, DRT hit rate and the number of radio tuning task instances completed were examined via the Pearson correlation coefficient (*r*), which is an index of the extent to which two variables have a linear relation. The correlations are presented in the following table.

Table 22. Correlation between Secondary Task and DRT Performance on Radio Tuning Trials

Variables	**Metric**	**Pearson r**	**p value**
Radio Tuning Instances	DRT RT	-0.56	< .0001
Radio Tuning Instances	DRT Hit Rate	0.54	< .0001
DRT RT	DRT Hit Rate	-0.69	< .0001

A moderate negative correlation was observed between the number of radio tuning task instances and the DRT response time. As the number of completed radio tuning task instances increased, the response time decreased. The direction of this effect does not reflect a tradeoff; rather it suggests that individuals who were able to work faster were also able to respond to DRT targets more quickly. A moderate positive correlation was observed between the number of radio tuning task instances and DRT Hit Rate. This also supports the interpretation that participants who were able to work faster on the radio tuning task were also able to detect more DRT targets. Finally, the DRT response time was negatively correlated with the DRT hit rate. The direction of this correlation is also consistent with the model of improved overall performance; participants who were able to detect more targets were also able to respond more quickly. Together, these results suggest that there is no evidence of a tradeoff among the participants between DRT and secondary task performance on radio tuning trials.

3.10 Parametric Versus Non-Parametric Test Comparison

The ISO working group has expressed concern about the use of normal parametric statistical tests on the hit rate metric, due to the expectation that the assumption of normality is seriously violated with this metric (ISO, 2014). However, because the response time metric is also known to violate this assumption, non-parametric statistics were used to analyze both metrics in this study. To explore the implications of using parametric versus non-parametric statistical tests, a subset of the comparisons presented in Table 18 was used to directly compare the two approaches. Specifically, the simulator venue data was selected for this purpose. The set of response time comparisons presented in Table 18 which summarized statistical test results for 2-minute and 1-minute data collection intervals with two independent samples of 24 participants was redone using (parametric) paired t-tests instead of the (non-parametric) Wilcoxon signed-rank tests. To ensure comparability with the non-parametric test results, the probability values were adjusted for familywise error in the same manner that the non-parametric test results were adjusted. The results are presented in Table 23. In this table, the comparisons of relevance are between the left and right halves of the table. Test outcomes that differ between parametric and non-parametric testing are highlighted. Analyses presented on the left side are identical to those presented previously.

Table 23. Parametric Versus Non-Parametric Statistical Test Outcomes by DRT, Data Collection Interval and Sample: Mean Response Time

	Non-Parametric (Wilcoxon) Results					Parametric (t-test) Results			
	2 minute		1 minute			2 minute		1 minute	
	Sample (N = 24)								
HDRT Mean RT	**1**	**2**	**1**	**2**		**1**	**2**	**1**	**2**
0-Back - 1-Back	<.0001	<.0001	0.0026	0.0002		0.0008	<.0001	0.0135	0.0004
0-Back - Radio Tune	<.0001	<.0001	<.0001	<.0001		<.0001	<.0001	<.0001	0.0001
0-Back - Baseline	0.0428	0.0921	0.1487	0.1327		0.0749	0.0743	0.1964	0.0957
1-Back - Radio Tune	0.0050	0.0447	0.0530	0.1327		0.0176	0.0312	0.0925	0.0953
TDRT Mean RT	**1**	**2**	**1**	**2**		**1**	**2**	**1**	**2**
0-Back - 1-Back	0.0014	<.0001	0.0497	0.0001		0.0088	<.0001	0.0687	0.0005
0-Back - Radio Tune	<.0001	<.0001	0.003	<.0001		<.0001	<.0001	0.0071	<.0001
0-Back - Baseline	0.0171	<.0001	0.0497	0.0015		0.0625	0.0003	0.0687	0.0013
1-Back - Radio Tune	0.0114	0.0171	0.0497	0.0051		0.0214	0.0192	0.1751	0.0069
RDRT Mean RT	**1**	**2**	**1**	**2**		**1**	**2**	**1**	**2**
0-Back - 1-Back	0.0007	<.0001	<.0001	<.0001		0.0037	<.0001	0.0022	0.0010
0-Back - Radio Tune	<.0001	<.0001	<.0001	<.0001		<.0001	<.0001	<.0001	<.0001
0-Back - Baseline	0.0534	0.1973	0.0534	0.0717		0.0313	0.1974	0.0257	0.1368
1-Back - Radio Tune	<.0001	<.0001	<.0001	0.0213		0.0001	0.0001	<.0001	0.0266

Of the 48 pairs of comparisons presented in this table, 7 (15%) revealed differences in statistical test outcome relative to the standard ($p < .05$) statistical criterion. All of the comparisons that exhibited different statistical outcomes had probability values that were close to the criterion value for either the parametric, non-parametric, or both results. When both probability values associated with a single discrepancy are close to the probability criterion boundary ($p = 0.05$), it is unlikely that the discrepancy has substantive meaning. Of the seven comparisons with discrepant test outcomes, five of the test outcomes had smaller probability values in the non-parametric testing versus two, which had smaller probability values in the parametric testing. The direction of these differences, while not strong, suggests that the non-parametric tests were at least as powerful as the parametric tests, assuming (as was done at the outset) that all of the four comparisons reflected effects that were different in reality. A similar comparison of statistical test probabilities using the hit rate metric is presented in Table 24.

Table 24. Parametric Versus Non-Parametric Statistical Test Outcomes by DRT, Data Collection Interval and Sample: Mean Hit Rate

	Non-Parametric Results					Parametric t-test Results			
	2 minute		1 minute			2 minute		1 minute	
	Sample (N = 24)								
HDRT Mean Hit Rate	**1**	**2**	**1**	**2**		**1**	**2**	**1**	**2**
0-Back - 1-Back	0.0042	0.0002	0.0344	0.0524		0.0107	0.0007	0.0425	0.0954
0-Back - Radio Tune	<.0001	<.0001	0.0008	<.0001		0.0003	<.0001	0.0011	0.0004
0-Back - Baseline	0.4579	0.0112	0.5723	0.0020		0.4966	0.0216	0.682	0.0178
1-Back - Radio Tune	0.0027	0.0006	0.0255	<.0001		0.0107	0.0011	0.0236	0.0011
TDRT Mean Hit Rate	**1**	**2**	**1**	**2**		**1**	**2**	**1**	**2**
0-Back - 1-Back	0.0005	<.0001	0.0005	0.0935		0.0021	0.0048	0.0007	0.0863
0-Back - Radio Tune	<.0001	<.0001	<.0001	<.0001		<.0001	<.0001	<.0001	0.0002
0-Back - Baseline	0.3630	0.1199	0.1762	0.2687		0.3414	0.0796	0.1303	0.1682
1-Back - Radio Tune	<.0001	0.0015	<.0001	0.0002		0.0003	0.0048	0.0004	0.0009
RDRT Mean Hit Rate	**1**	**2**	**1**	**2**		**1**	**2**	**1**	**2**
0-Back - 1-Back	0.0510	0.0005	0.0234	0.0027		0.0402	0.0050	0.0258	0.0044
0-Back - Radio Tune	<.0001	<.0001	<.0001	0.0006		0.0014	<.0001	0.0007	0.0014
0-Back - Baseline	0.3397	0.0552	0.7930	0.0396		0.5930	0.0745	0.7268	0.0449
1-Back - Radio Tune	0.0047	0.0252	0.0029	0.0396		0.0162	0.0344	0.0079	0.0449

As shown in the highlighting, only 1 of 48 pairs of comparisons provided discrepant statistical test outcomes for hit rate. The probability values associated with the single discrepancy are both so close to the probability criterion boundary ($p = 0.05$) that it is highly unlikely that the discrepancy has substantive meaning. Therefore, the non-parametric and parametric statistical tests provide essentially identical results for this set of comparisons. This finding supports the conclusion that the expected violation of underlying assumptions of the parametric tests had no effect on the test outcomes.

The following two figures present the underlying distribution of response time (Figure 58) and hit rate (Figure 59) values together with normality templates to demonstrate the extent to which the distributions deviate from normality. Both distributions reveal deviations from normality as expected.

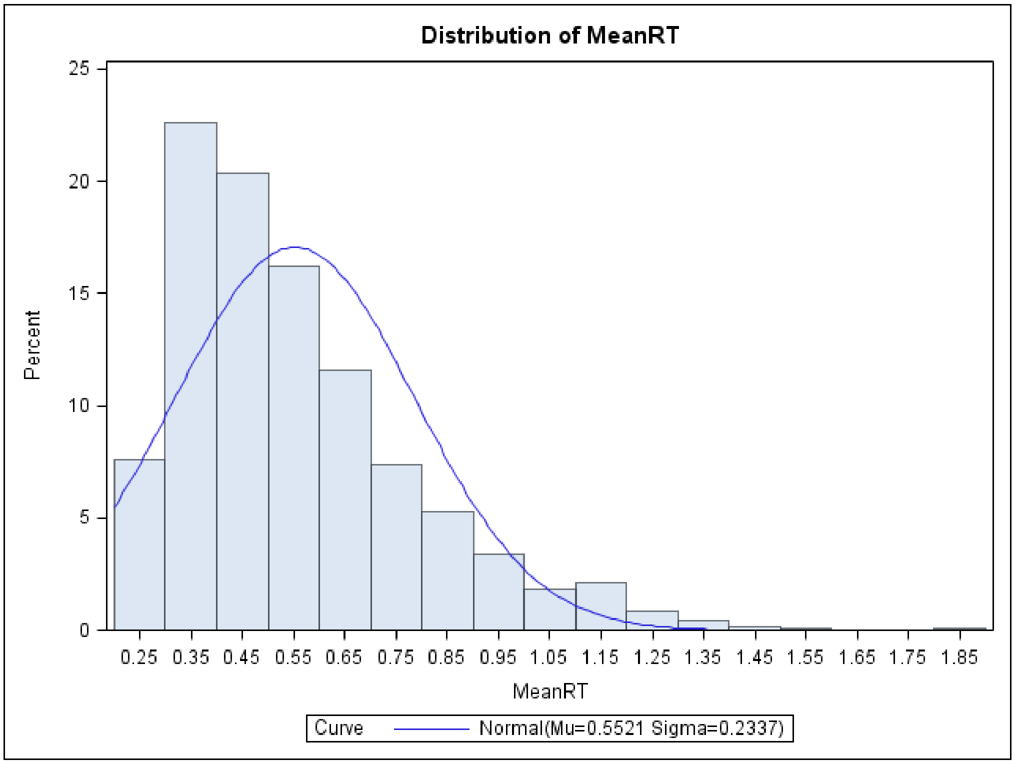


Figure 58. Distribution of Mean Response Times With Normal Template

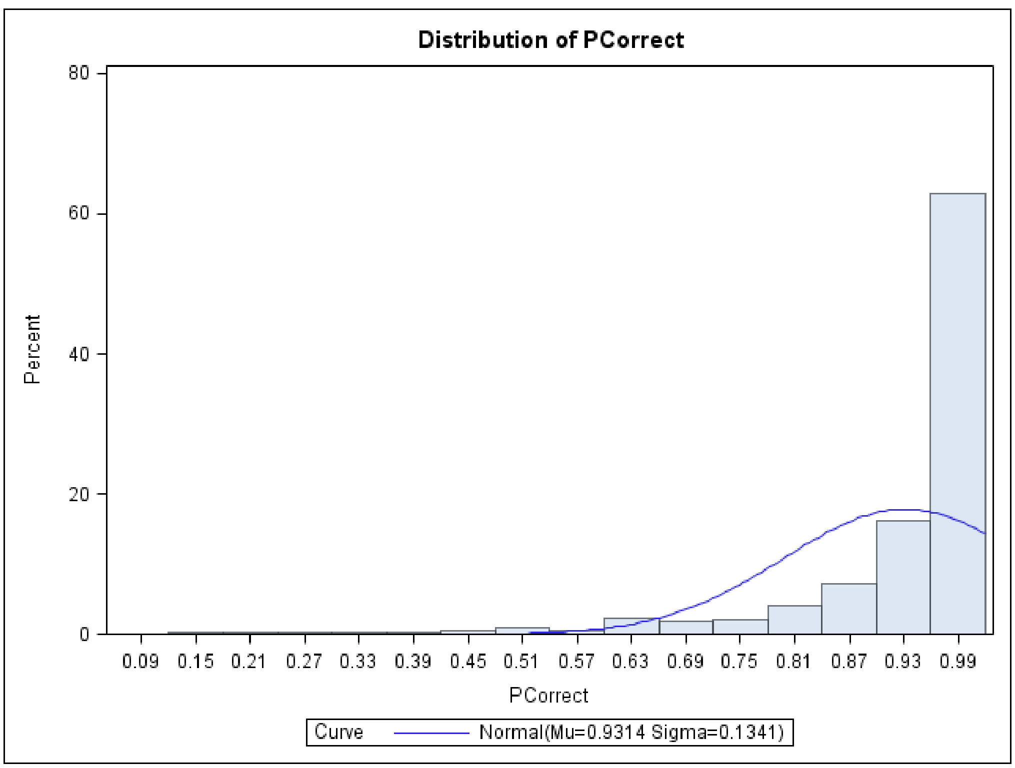


Figure 59. Distribution of Hit Rate Values With Normal Template

3.11 Summary of Findings

Key Findings

The main objectives of the study were embodied in three questions presented in Section 3.5. In this section, the results that provide relevant information in relation to these questions are summarized. Many of the analyses refer to the following four planned comparisons:

1. 0-Back versus 1-Back
2. 0-Back versus Radio Tuning
3. 0-Back versus Baseline
4. 1-Back versus Radio Tuning

Comparisons 1 and 3 were diagnostic comparisons used for assessing the sensitivity of the DRT variants. Comparisons 2 and 4 were used for benchmarking purposes.

Question 1. Do differences exist among DRT conditions (HDRT, TDRT and RDRT) that would make one preferable for use in testing?

The following findings are based on the full sample (N = 48) analyses:

- For response time, all DRT variants were able to detect all four targeted differences with minor exceptions in both test venues.

Comparison 3 was smallest among the targeted differences, thus representing the most difficult test of metric sensitivity. The following findings focus on this specific comparison:

- For response time in the simulator, TDRT was most sensitive to the difference in Comparison 3.
- For response time in the non-driving venue, all DRT variants were able to detect the difference in Comparison 3.
- For hit rate, none of the DRT variants was consistently able to detect the Comparison 3 difference in either test venue.

The following findings are based on the small sample (N = 24) analyses:

- For both metrics in the driving simulator, both HDRT and TDRT successfully detected differences associated with the two diagnostic comparisons (1 and 3); RDRT was weaker.
- For both metrics in the driving simulator, the TDRT exhibited the highest level of agreement (test-retest reliability) for a 2-minute data collection interval.
- Effect sizes, both raw and standardized, revealed evidence suggesting that unlike TDRT and HDRT, RDRT measures different components of distraction when used to assess tasks performed with different interfaces.

Question 2. Do differences exist between test venues (Simulator, Non-Driving) that would make one preferable for use in testing?

- Hit rate differences were weaker in the non-driving venue; consistently high hit rates, reflecting ceiling effects due to consistently near-perfect target, reduced sensitivity of hit rate in this venue.
- In the non-driving venue, differences between secondary task conditions were smaller for hit rate than for response time.
- For response time, there was more sensitivity among conditions in the non-driving venue than in the simulator venue.
- For response time, ES values were consistently larger in the non-driving venue relative to the simulator venue.

Question 3. Do differences exist among data collection intervals of different durations?

The following findings are based on the full sample (N = 48) analyses:

- All effects were detected using both 2- and 3-minute data collection intervals.
- Test outcomes using 1-minute data collection intervals were consistent with longer interval results for response time, but weaker for hit rate.
- Test outcomes using 30-second intervals were weaker for both response time and hit rate than for longer intervals.
- For response time in the simulator, the robustness of the Comparison 3 difference was slightly greater using the 2-minute data collection interval relative to the 1-minute interval.
- For response time in the non-driving venue, the Comparison 3 difference was detected robustly and consistently using both 1- and 2-minute data collection intervals.
- For hit rate in both test venues, the Comparison 3 difference could not be detected at either 1- or 2-minute data collection intervals.

The following findings are based on the small sample (N = 24) analyses:

- For both metrics in the simulator, test outcomes based on a 1-minute data collection interval were worse for HDRT and TDRT, but slightly better for RDRT relative to those observed with a 2-minute interval.
- For both metrics in the non-driving venue, test outcomes based on a 1-minute data collection interval were slightly worse for all DRT variants relative to those observed with a 2-minute interval.
- In the simulator, among all DRT variants at 1- and 2-minute intervals, TDRT exhibited highest test-retest reliability at the 2-minute interval.
- In the non-driving venue, there were no differences among DRT variants in agreement between samples using both 1- and 2- minute data collection intervals.
- For both test venues, effect sizes were smaller at shorter data collection intervals.
- For response time data, issues related to the lack of normality are eliminated based on the Central Limit Theorem for data collection intervals (2+ minutes) that allow trial means to be computed using data from approximately 30 individual trials.

Other results, not directly related to the 3 key questions are summarized in the following sections.

General Results

- DRT response times obtained in the non-driving venue were consistently faster than those obtained in the driving simulator venue. Hit rate values were consistently higher in the non-driving venue than in the driving simulator venue. These patterns reflect the reduced difficulty of target detection when there was no concurrent driving task.
- Both DRT response time and hit rate means showed increasing performance degradation (longer response times and lower hit rates) relative to the baseline condition among secondary task conditions in the following order: 0-back, 1-back and radio tuning.
- Analyses performed on aggregated response time data revealed consistent results across data collection intervals of 1 minute, 2 minutes and 30 seconds.
- Tactile condition (TDRT) response times were significantly longer than those in the two visual DRT conditions.
- Analyses performed on aggregated hit rate data revealed differences between DRT by test venue combinations due primarily to the consistently higher hit rates observed in the non-driving venue. As a result, hit rate differences between DRT conditions and secondary task conditions were more pronounced in the driving simulator venue.

Trends Across Time

- Response time means did not remain consistent across successive 30-second segments of the data collection interval for some DRT by test venue combinations; however, the differences had only minor effects on statistical test outcomes related to differences between secondary task conditions.
- Response time means did not remain consistent across successive 15-second segments of the data collection interval. These differences were more pronounced than those observed with 30-second segments and had more effects on statistical test outcomes.
- Differences across successive 15- and 30-second segments of the data collection interval were more evident and larger in the driving simulator venue than in the non-driving venue.
- Mean hit rate values based on 30-second segments were generally less variable across time than response time means. Hit rate values based on 15-second segments exhibited a slight increase in variability relative to the 30-second segment hit rates.

Secondary Analyses

- The failure to control head position in the RDRT condition affected response time differences between visual manual (radio tuning) and auditory vocal (0-back) conditions in ways that were not apparent in the other DRT conditions.
- Participants who completed more secondary task instances in a fixed data collection interval were also able to detect more DRT targets and respond more quickly to those targets than participants who completed fewer secondary task instances. This pattern suggests that participants were not engaged in strategies involving (speed/accuracy) tradeoffs that could have contaminated test results.

- Based on the family of planned comparisons using data from 1- and 2-minute data collection intervals, there were no systematic differences in test outcome between parametric and non-parametric statistical tests for both response time and hit rate. Differences between independent samples were more pervasive than differences between statistical approaches. There was no loss of power associated with parametric tests.

4. DISCUSSION

4.1 Key Questions

The study objectives were embodied in three key questions. This section will integrate the relevant findings and provide answers to these questions to the extent possible.

1. Do differences exist among DRT variants (HDRT, TDRT and RDRT) that would make one preferable for use in testing?

Most generally, the results of the present study revealed only minor differences among the DRT variants. All three DRT variants were generally successful in detecting the differences defined at the outset of the study. Results based on the response time metric were consistently positive with all DRT variants in both test venues. TDRT revealed some minor advantages in the driving simulator, including better sensitivity for detecting the most challenging of the targeted differences and slightly better test-retest reliability. None of the DRT variants stood out in the non-driving venue.

ES analyses revealed evidence suggesting that RDRT response time values may differ from those obtained with the HDRT and TDRT. In particular, elevated RDRT response times in the radio tuning condition appear to include variability due to conflicts between the RDRT and secondary task requirements when used with visual-manual tasks. HDRT and TDRT variants had smaller ES values for these differences, consistent with the effective operation of controls incorporated to minimize contributions from these variability sources. The implications of this potential problem are explained in greater detail below.

2. Do differences exist between test venues (simulator, non-driving) that would make one preferable for testing?

The main difference between venues had to do with hit rate, which was consistently less sensitive to the targeted differences in the non-driving venue than in the driving simulator venue. Hit rates were higher and approached the perfect-performance ceiling in the non-driving venue to the point that this metric was not particularly useful in this venue. The weakness of the hit rate metric in the non-driving venue was apparent both in statistical tests performed using aggregated data collapsed over DRT variants as well as in statistical tests performed using planned comparisons for each combination of DRT and test venue.

The response time metric in the non-driving venue was slightly more sensitive in detecting the targeted differences than in the driving simulator venue. This trend was reflected in stronger statistical outcomes and correspondingly larger ES values in the non-driving venue relative to the driving simulator venue.

These results raise the question of whether the hit rate metric is essential to the analysis of DRT performance. If DRT testing requires a sensitive measure of hit rate, then the present results suggest that the non-driving test venue may be unsuitable for this purpose. In contrast, if hit rate is not essential, then the increased sensitivity and easier implementation offered by the non-driving venue would represent significant advantages over the driving simulator venue. The question of whether both response time and hit rate metrics are needed is considered in greater detail in the next section.

3. Do differences exist among data collection intervals of different durations?

A set of omnibus analyses performed on aggregated response time data at collection intervals of 30 seconds, 1 minute, and 2 minutes found generally consistent patterns of differences, however larger ES values were associated with longer data collection intervals, with the biggest difference being observed between 2- and 1-minute intervals. The same pattern was observed for hit rate. These results suggest a progressive loss of sensitivity among the metrics at shorter data collection intervals.

Results based on planned comparisons with the small samples, provided support for using a 2-minute data collection interval. Focusing on the two diagnostic planned comparisons (1 and 3) for simulator testing, the 2-minute interval performed consistently better than the 1-minute interval. Overall, the combination of TDRT and a 2-minute data collection interval provided the best performance in the simulator venue. Among all DRT variants, the 2-minute data collection interval revealed a consistent advantage over the 1-minute interval; however, this advantage was considerably smaller in the non-driving venue than in the driving simulator venue.

4.2 Secondary Questions

The results of the study raised additional questions. This section will provide answers to the extent possible:

1. Does testing require one or two performance metrics?

The draft ISO standard supports using both response time and hit rate metrics in DRT testing, however, response time is the primary metric. Hit rate has been used both as a performance metric and as a quality indicator, whereby a specified accuracy criterion (e.g., 70% successful detection) must be attained on each trial to allow inclusion of the respective response time means in the analysis (Victor, Engström, & Harbluk, 2009).

There are additional reasons supporting the need for having valid measures of both response time and hit rate. The first reason concerns the ability to detect participants' use of different strategies in responding to DRT targets. If some participants decide to trade accuracy for speed while others strive for perfect detection, the response times may not be comparable across participants and the resulting differences in mean response times may not be valid (Victor, Engström, & Harbluk, 2009). A speed-accuracy tradeoff would be revealed by a positive correlation between the two metrics, reflecting faster response times together with lower hit rate values. Acceptable DRT performance is characterized by a negative correlation between the two metrics, reflecting the tendency of more demanding secondary tasks to elicit both lower hit rates and slower response times. The consistently high hit rates observed in the non-driving test venue suggest that participants did not trade accuracy for speed in that venue. However, if the lower hit rates observed in the driving simulator venue can be taken to suggest that a speed-accuracy tradeoff is more likely in that venue then there may be reason to consider the question of whether the two venues differ fundamentally in their validity in representing the demands of secondary task performance while driving. This question is addressed separately, below.

A second possible reason supporting the need for two valid metrics derives from the potential conflict between DRT responses and secondary task responses for tasks that require numerous and/or frequent button presses, for example, continuous scrolling accomplished using a push-button input that does not necessarily require diversion of vision away from the forward roadway view. Here, there is concern that the response requirements of the secondary task may conflict with the DRT response requirements thus slowing the DRT response time. In this situation, the longer response time values could be interpreted to suggest, perhaps inappropriately, that such tasks have higher attentional demands than tasks with greater demands on working memory but no significant response conflict. In this situation, one might expect the hit rate values among a set of secondary tasks to exhibit a different pattern than the corresponding response time metrics. Whether the two metrics could be used together to identify the occurrence of this conflict is an open question in need of additional theoretical and experimental work.

Having two valid measures of DRT performance provides redundancy and allows corroboration of findings. The present results suggest that the hit rate metric will not provide such information in the non-driving venue. However, the non-driving venue hit rate values may be suitable for purposes of quality assessment and to identify potential methodological issues such as a speed-accuracy tradeoff.

2. Does testing require a venue with demonstrated validity to the demands of driving?

If one accepts the premise that the driving simulator venue is a better representation of the demands of performing secondary tasks while driving than the non-driving venue, then the present experiment can be seen as offering an opportunity to validate the non-driving test venue relative to the driving simulator venue. In this regard, the pattern of results is generally consistent across test venues for response time, but not for hit rate. The ordering of response time means across secondary task conditions was consistent across the two venues. In particular, the consistent differences among the secondary task conditions in both venues indicate that the non-driving venue provides relative validity while the consistently smaller response time means in the non-driving venue reflect a lack of absolute validity. In contrast, the consistently high hit rates in the non-driving venue created a ceiling effect that effectively attenuated the differences between secondary task conditions resulting in a lack of both relative and absolute validity. Although not surprising due to the absence of a concurrent driving task, the results support the conclusion that the non-driving test venue does not provide a valid representation of the demands of secondary task performance while driving. The non-driving venue can be expected to provide valid information concerning the relative ordering of response times among a set of secondary tasks; however, the hit rate information does not appear to be valid in the non-driving venue.

The driving simulator combines secondary task performance with a simple driving task and thus better represents the demands of secondary task performance while driving than the non-driving venue. Whether the relative validity offered by the non-driving venue is sufficient for testing will depend on the structure of the test protocol. For example, if DRT response time values are to be compared to an absolute criterion that has real-world meaning, then the relative validity of the non-driving venue would not be appropriate. Response time values obtained in the driving simulator may also be inappropriate if they are shown to be consistently faster than those obtained in on-road driving (e.g., Strayer et al., 2013). To support this test approach, the response time values obtained in the driving simulator would need to be compared to those obtained in on-road driving. However, if the protocol involves comparing tasks with a standardized calibration task that can be included in each test (e.g., 1-back), then both test venues would be expected to provide appropriate information using the response time metric.

3. Do all three DRT variants provide consistent information about attentional distraction?

Based on theoretical considerations and recent findings (e.g., Merat & Jamson, 2008; ISO, 2014), the three DRT variants are expected to provide comparable information when used to assess tasks performed with auditory-vocal interfaces; however, they are expected to provide different (and potentially not comparable) information when used to assess tasks performed with visual-manual interfaces. For DRT variants that use visual targets, there are two possible sources of response delay associated with visual-manual tasks. The first source of delay is due to the conflict between visual requirements of the secondary task and those of the DRT. Responses to visual DRT targets presented while drivers are processing information on an in-vehicle visual display can be delayed due to the conflict in demand for visual resources, even if the target is visible when presented. Both HDRT and RDRT incorporate this potential conflict. The second source of delay is due to differences in the participant's head position when the target is presented. If the participant is looking inside the vehicle when a target is presented remotely in the forward field of view, the target may not be visible until the point of gaze returns to the forward roadway. Because the HDRT controls the position of the target relative to the participant's head, the potential for delay due to this problem is mostly eliminated, assuming that the HDRT target is visible at all times. Thus, the potential for response delay due to this conflict is associated primarily with the RDRT.

According to this model, response time delays attributable to these conflicts should be apparent in visual-manual tasks that require drivers to acquire task-relevant information visually, but not for auditory-vocal tasks like n-back that have no visual requirements. To the extent that delays associated with these potential conflicts are additive, the cumulative delay associated with the RDRT should be greater than that associated with the HDRT. To date, this model has not been widely tested. The present experiment was not designed to isolate this effect; however, the results provided relevant, if only suggestive, evidence. Using simulator data from the 3-minute data collection interval, differences in mean response time between radio tuning (attentional plus visual-manual components) and 1-back (attentional with no visual-manual components) conditions were computed. The difference was greatest in the RDRT condition (.77-.62 = .15), which is consistent with the hypothesized contribution of both visual conflicts. The difference was smallest in the TDRT condition (.82-.75 = .07), which is expected due to the absence of either potential visual conflict. Finally, the difference was intermediate in the HDRT condition (.75-.65 = .10), which is consistent with the operation of one but not both potential conflicts, under the assumption of additivity. This effect was not observed in the simulator hit rate means, nor did it appear for either metric in the non-driving venue. However, the pattern was reflected in elevated ES values in the simulator response time analyses that were computed using Cohen's d. In the non-driving venue, the pattern of elevated differences between tasks performed with different interfaces was not apparent among the raw mean differences; however, the standardized ES values (Cohen's d) did reveal this effect. Interpretation of the differences in the non-driving venue was complicated by the fact that a similar pattern (elevated differences in RDRT) was observed for comparisons between tasks that did not involve visual-manual interfaces. The results were therefore not unequivocal in their support for this phenomenon in the non-driving venue. While these comparisons provide preliminary support for the hypothesized operation of the two visual conflicts, a more direct comparison is needed, ideally in which individual response time values are related to drivers' head position and hand activity at the time of visual target onset. The results do support the conclusion that the RDRT provides slightly different information than the other two DRT variants.

4. Is it necessary that the DRT be a pure measure of attentional distraction?

Generally, there are two perspectives that support different answers to this question. From a basic scientific perspective, it is generally always better to have metrics that measure the same behavioral constructs in all conditions. Without such integrity, it is very difficult to interpret comparisons across tasks that may involve combinations of different underlying behavioral mechanisms. The practical benefit in the context of DRT testing is that precision in identifying the locus of a problem would allow developers to better determine exactly what changes to an interface would be necessary to influence the metric value. In contrast, overlap in the underlying psychological constructs represented in the metrics, as suggested above for the RDRT and HDRT metrics, would compromise the metrics' independence and could create uncertainty about how to address problems underlying non-conforming test outcomes. Moreover, to the extent that DRT metrics represent an unknown mixture of visual-manual and attentional distraction effects, it could be difficult to interpret differences across tasks performed with different interfaces using these metrics because the underlying scale could not be assumed to have the property of equal intervals. DRT metrics that emphasize isolation of behavioral constructs eliminate this problem and provide a stronger basis for comparisons of tasks performed with different interfaces.

The second perspective reflects an overriding interest in relating the metric values directly to crash risk. If a DRT metric has an established and direct relation to crash risk, then one could argue that the underlying behavioral mechanisms that contribute to the metric are less important than the connection to safety. This perspective may also reflect the practical reality that it is very difficult to identify and separate the underlying behavioral contributions to distracted driving. Moreover, if contemporary tasks are more likely than earlier generations of tasks to involve combinations of visual-manual and auditory-vocal components, then assessing these tasks may be better served by measures that assess distraction potential comprehensively.

Distraction is generally considered to be a multidimensional phenomenon, such that a comprehensive assessment will require a number of different metrics. NHTSA's Phase 1 Guidelines focus on the visual components of distraction. The DRT variants considered in the present study offer different possibilities including a relatively pure measure of attentional distraction (TDRT) and a mixture of attentional and visual-manual components that differs depending on the specific task demands (RDRT). Neither the visual nor the DRT metrics provide a comprehensive assessment of distraction potential. Therefore, unless DRT metrics can be shown to have a direct relation to crash risk, they will be more useful if they always measure the same underlying behavioral construct. The question of how DRT performance relates to safety is addressed next.

5. What is the relation between DRT performance and safety?

Distracted drivers are at increased risk of failing to notice and/or responding too slowly to an emerging hazardous situation. Two experimental methods have been used to assess the drivers' slowed responses. Object and event detection (OED) methods involve recording drivers' responses to unexpected hazards, such as a stopped vehicle in the roadway ahead (e.g., Lee, McGehee, Brown, & Reyes, 2002; Strayer, Drews, & Johnson, 2003). DRT methods record drivers' responses to simple frequently-presented targets. Recent theoretical advances (e.g., Engström, 2010) have clarified the behavioral mechanisms underlying DRT performance, providing a basis for differentiating DRT target detection from OED methods. In particular, Engström (2010) has argued that the use of frequently occurring and expected targets in DRT methods engages different brain mechanisms than those required to respond to the unexpected infrequently occurring OED targets. This analysis implies that from the perspective of underlying brain mechanisms, DRT methods may not be valid representations of OED trials. If this is true, the connection between DRT metrics and crash risk would be weaker than the connection between OED metrics and crash risk because the OED trials better represent distracted-driving crash situations. However, there is no consistent empirical support for the hypothesized differentiation between the two methods based on brain mechanisms. Rather, one recent study provides modest support for the conclusion that brake response time and DRT response time are highly correlated, when used to assess a set of voice-based tasks (Strayer, Cooper, Turrill, Coleman, Medeiros-Ward, & Biondi, 2013). Moreover, arguments based on face validity have enhanced the credibility and facilitated the widespread use of the DRT and its predecessors. Merat and Jamson (2008) argued that using visual signals in DRT tasks is more ecologically valid than targets presented to different modalities. Among those considered in the present study, the RDRT most resembles the earlier versions and thus has the highest level of face validity, due primarily to the remote target location. In contrast, the newer variants, particularly the TDRT and HDRT, have traded face validity for better control of potential visual conflicts and elimination of unwanted variability, which are intended to improve metric sensitivity.

Although OED methods may have a stronger and more direct connection to drivers' responses in critical real-world driving situations, there are practical and methodological considerations that favor the use of DRT methods for assessing attentional distraction. First, DRT methods and their predecessors have a considerably stronger history of demonstrated sensitivity to attentional distraction than OED methods (Victor, Engström, & Harbluk, 2009). Based on this work, there is emerging consensus that DRTs are sensitive to the attentional demands of secondary tasks. Second, the use of surprise events in OED methods severely limits the number of presentations that can be made and thus the usefulness of this approach for testing protocols that require comparisons across multiple conditions. Third, the fact that DRT response times represent data collected from on average 15 trials per minute, versus a single response in OED methods, provides increased stability in the metric and better adherence to the underlying assumptions required for statistical testing. The strong empirical foundation together with the practical advantages of DRT methods led to the ISO decision to select the DRT as the most promising method for assessing differences in the potential for distraction due to differences in attentional demand between tasks (ISO, 2013).

DRT research has shown that patterns of results are consistent across target modality (e.g., Merat & Jamson, 2008), which supports the conclusion that all DRT variants measure the effects of secondary task performance on drivers' attention, independent of target modality. It follows that a delay in responding to a tactile target would be expected to have the same effect on crash likelihood as a similar delay in responding to a visual target. Additional experimentation will be necessary to better determine the relation between DRT and OED response time metrics.

6. What additional information is needed to develop a DRT test protocol?

The results showed that any of the DRT variants in either test venue could reliably detect the difference that had been established, based partly on ISO work, as the minimum requirement for a DRT, namely the difference between 0-back and 1-back (Comparison 1). In search of a more challenging test of metric sensitivity, attention was then directed at the difference between 0-back and Baseline driving (Comparison 3), which although smaller in magnitude, is considered meaningful based on differences observed in physiological and driving performance metrics (Mehler, Reimer, & Coughlin, 2012). Test outcomes associated with this comparison did help differentiate among the DRT variants in terms of sensitivity.

What is missing at this point is a consensus concerning the required sensitivity of the DRT metrics. In other words, what size "dose" of attentional distraction represents a meaningful difference that should be detectable by a metric? Is it the larger dose represented by Comparison 1, the smaller dose represented by Comparison 3, or a different amount altogether? This question leads naturally to the following 2 questions:

- What is the maximum acceptable dose of attentional demand?
- How much additional attentional demand represents an unacceptable level?

These are not empirical questions that can be addressed via experimentation. Answering them will require reasoned judgment. But experimentation can provide information that will help answer these questions. In particular, what is needed is DRT performance data obtained from a set of real world tasks performed with auditory-vocal interfaces with levels of attentional demand that range from clearly acceptable to clearly unacceptable.

Among visual-manual tasks, consensus emerged over time that entering a street address involved too much distraction. Text messaging also was clearly unacceptable. These two tasks provided useful anchors in helping to establish a boundary between acceptable and unacceptable levels of visual-manual distraction. At this point, there is no clear consensus auditory-vocal task that represents either the maximum acceptable or an unacceptable level of attentional distraction. The n-back task, which comprises the essential elements of a paced conversation (listening, remembering, and speaking) over a wide range of attentional demands, can provide help in defining anchors on a hypothetical scale of attentional distraction. Although the relation between various n-back demand levels and crash risk has not been established, experimental results based primarily on subjective ratings have indicated that the 1-back represents a level of demand that is generally acceptable while driving. In contrast, the 2-back represents an unacceptable level of attentional demand while driving (Reimer, Mehler, Dobres, & Coughlin, 2013; Ranney, Baldwin, Parmer, Domeyer, Martin, & Mazzae, 2011). When placed on a hypothetical continuum, these points define the boundaries of a critical region, which will include the as yet undetermined point that represents the maximum acceptable level of attentional demand. Obtaining comparable DRT performance data for a number of real-world auditory-vocal tasks will help populate the hypothetical continuum and narrow the critical region until the maximum acceptable level emerges.

4. CONCLUSIONS

The results support the following conclusions:

- Differences among DRT variants were small, but the TDRT was slightly more sensitive than the HDRT and RDRT when used in the driving simulator venue.
- The three DRT variants provided comparable sensitivity in the non-driving venue.
- The hit rate metric was generally less sensitive than response time in both test venues; consistently near-perfect target detection significantly reduced the sensitivity of this metric in the non-driving venue.
- The response time metric revealed slightly greater sensitivity in the non-driving venue than in the driving simulator venue.
- A 2-minute data collection interval provided optimal sensitivity for driving simulator venue testing, particularly for small-sample comparisons. A slightly shorter interval may be feasible in the non-driving venue without significant loss of metric sensitivity.
- The non-driving venue offers easier implementation plus slightly greater sensitivity for the response time metric. The hit rate metric provided insufficient sensitivity for consistently discriminating among secondary task conditions in this venue.
- The driving simulator venue provides a more valid representation of the concurrent demands of driving and secondary task performance. Both response time and hit rate metrics provided adequate sensitivity for discriminating among secondary task conditions in this venue.

5. REFERENCES

Alliance of Automobile Manufacturers (2006). *Statement of principles, criteria and verification procedures on driver interactions with advanced in-vehicle information and communication systems*. Washington, DC: Alliance of Automobile Manufacturers, June 26.

Canty, A., & Ripley B. (2014). Boot: Bootstrap R (S-Plus) Functions. R package version 1.3-11.

Caird, J. K., Willness, C. R., Steel, P., & Scialfa, C (2008). A meta-analysis of the effects of cell phones on driver performance. *Accident Analysis and Prevention*, *40*, 1282-1293.

Cumming, G. (2014). The new statistics: Why and how. *Psychological Science, 25*,7-29.

Cumming, G. (2013). *Understanding the new statistics: Effect sizes, confidence intervals, and meta-analysis*. New York, NY: Routledge.

Davison, A. C., & Hinkley, D. V. (1997). *Bootstrap methods and their applications*. Cambridge University Press, Cambridge. ISBN 0-521-57391-2

Devore, J. L. (2008). *Probability and statistics for engineering and the sciences* (7th Ed.).Belmont, CA: Thomson Brooks/Cole.

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, pp. 90-103.

Harms, L., & Patten, C. (2003). Peripheral detection as a measure of driver distraction. A study of memory-based versus system-based navigation in a built-up area. *Transportation Research Part F, 6*, 23-36.

Horrey, W. J., & Wickens, C. D. (2006). Examining the impact of cell phone conversations on driving using meta-analytic techniques. *Human Factors*, *48*, 196-205

Horst, A. R., van der, & Martens, M. H. (2010). The peripheral detection task (PDT): On-line measurement of driver cognitive workload and selection attention. In G. Rupp (Ed.), *Performance metrics for assessing driver distraction: The quest for improved road safety*. Warrendale, PA: SAE International, pp. 73-89.

ISO CD 17488 (2013). Road Vehicles - Transport information and control systems - Detection Response Task (DRT) for assessing selective attention in driving. Committee Draft, ISO TC 22/SC13/WG8.

Kelley, K. (2005). The effects of nonnormal distributions on confidence intervals around the standardized mean difference: Bootstrap and parametric confidence intervals. *Educational and Psychological Measurement., 65*, 51-69.

Kelley, K., & Preacher, K. J. (2012). On effect size. *Psychological Methods, 2012,17*,137-152.

Mehler, B., Reimer, B., & Coughlin (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,* Vol.*54*, no. 3, pp. 396-412.

Mehler, B., Reimer, B., & Dusek, J.a. (2013). MIT AgeLab delayed digit recall task (n-back). MIT AgeLab White Paper Number 2013-4. Cambridge, MA: Massachusetts Institute of Technology,.

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

Miura, T. (1986). Coping with situational demands: A study of eye movements and peripheral vision performance. In A. G. Gale, M. Freeman, C. M. Haslegrave, P. Smith, & S. Taylor (Eds.), *Vision in vehicles I* Amsterdam: Elsevier.

Myers, J. L, & Well, A .D. (1991). *Research design and statistical analysis*. New York: Harper Collins.

National Highway Traffic Safety Administration. (2013, April 26). *Visual-Manual NHTSA Driver Distraction Guidelines for In-Vehicle Electronic Devices.* (Public Notice. Document ID: NHTSA-2010-0053-0135. Federal Register Number: 2013-09883). Retrieved from eRegulations.gov at http://federal.eregulations.us/rulemaking/document/NHTSA-2010-0053-0135

Pettitt, M., Burnett, G., & Stevens, A. (2005). *Defining driver distraction*. Paper presented at World Congress on Intelligent Transport Systems, San Francisco, CA.

R Core Team (2014). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. URL http://www.R-project.org/.

Ranney, T. A. (2008, April). Driver distraction: A review of the current state-of-knowledge. (Report No. DOT HS 810 787). Washington, DC: National Highway Traffic Safety Administration. Available at www.nhtsa.gov/DOT/NHTSA/NRD/Multimedia/PDFs/Crash%20Avoidance/2008/810787.pdf.

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.* (Report No. DOT HS 811 463). Washington, DC: National Highway Traffic Safety Administration.

Ranney, T. A., Baldwin, G. H. S., Parmer, E., Martin, J., & Mazzae, E. N. (2011). *Distraction effects of manual number and text entry while driving.* (Report No. DOT HS 811 510). Washington, DC: National Highway Traffic Safety Administration.

Ranney, T. A., Harbluk, J. L., Smith, L., Huener, K., Parmer, E., & Barickman, F. (2003). *The effects of voice technology on test track driving performance: Implications for driver distraction.* (Report No. DOT HS 809 525). Washington, DC: National Highway Traffic Safety Administration.

Ranney, T. A., Mazzae, E. N., Baldwin, G. H. S., & Salaani, M. K. (2007). *Characteristics of voice-based interfaces for in-vehicle systems and their effects on driving performance.* (Report No. DOT HS 810 867) Washington, DC: National Highway Traffic Safety Administration.

Reimer, B., Mehler, B., Dobres, J. U., & Coughlin, J. F. (2013). *The effects of a production level "voice-command" interface on driver behavior: Reported workload, physiology, visual attention, and driving performance.* (Technical Report 2013-17A). Cambridge, MA: Massachusetts Institute of Technology Age Lab.

Strayer, D. L., Cooper, J. M., Turrill, J., Coleman, J., Medeiros-Ward, N., & Biondi, F. (2013). *Measuring cognitive distraction in the automobile*. Washington DC: AAA Foundation for Traffic Safety.

US-EU Bilateral ITS Technical Task Force, European Commission Expert Focus Group on Driver Distraction: Definition and Research Needs. (2010). *Digital agenda for Europe: A Europe 2020 initiative*. Retrieved from http://ec.europa.eu/digital-agenda/en/news/september-2011-driver-distraction-definition-and-research-needs-report-expert-focus-group-us%E2%80%90eu.

Victor, T. W., Engström, J., & Harbluk, J. L. (2009). Distraction assessment methods based on visual behavior and event detection. In M. A. Regan, J. D. Lee, & K. L. Young (Eds.),*Driver distraction: Theory, effects and mitigation* (pp. 135-165). Boca Raton, FL: CRC Press.

Westfall, P. H., Tobias, R. D., & Wolfinger, R. D. (2011). *Multiple comparisons and multiple tests using SAS* (2nd Ed.). Cary, NC: SAS Institute Inc.

Appendix A: Recruitment Materials

The recruitment materials are presented in this appendix in the order of their use. A prospective participant responds to a Recruitment Advertisement by proceeding online to an Application Link, which provides study overview information and directs the individual to a secure Contact Information Form. Individuals successfully submitting the Contact Information Form and meeting the needed criteria are sent an email containing a link to the secure Screening Form. Individuals who successfully complete the Screening Form and are deemed eligible for participation are contacted for appointments as needed to fill the test matrix.

Recruitment Advertisement

PARTICIPANTS NEEDED FOR DRIVING STUDY

Receive $42 per hour, plus mileage allowance, for 7 to 8 hours of participation

We are seeking participants for a simulator study of driving performance

The study will be conducted by:
Transportation Research Center Inc. for the National Highway Traffic Safety Administration (NHTSA) of the U.S. D.O.T.
At the proving ground in East Liberty, Ohio
Weekday Sessions

Participation Requirements:
* Must be 18 to 70 years old
* Good general health
* Must have a valid U.S. driver's license with no restrictions other than corrective lenses
* Drive at least 3000 miles per year
* Must have experience using a cell phone while driving

We are especially in need of drivers 55 to 70 years old.

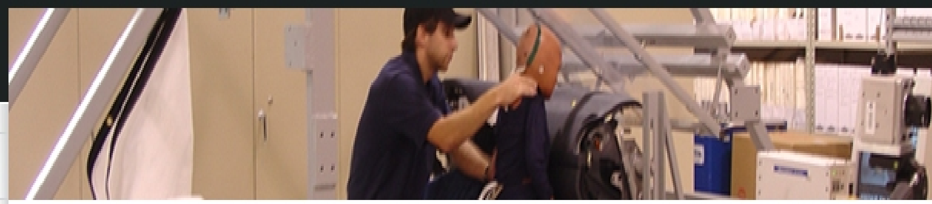
To apply, go to: http://www.trcpg.com/about-trc/research-study.aspx
and click on the link for the online application form.

Application Link

TRC Transportation Research Center Inc®

HOME | ABOUT TRC | FACILITY TOUR | OUR SERVICES | EMPLOYMENT | CONTACT US

SEARCH:



- History of TRC
- Annual Report
- Company Officers
- Community Relations
- Maps to TRC
- Local Hotels
- Company Survey
- Registrations
- Accreditations
- Vision & Mission
- PACE Awards
- Industry Links
- Research Study

Research Study

The Transportation Research Center Inc. is conducting a research study for the United States Department of Transportation's National Highway Traffic Safety Administration (NHTSA). The study will evaluate different tools that researchers use to measure distraction caused by in-vehicle technologies and portable devices including cell phones.

Participation involves one session of approximately 7-8 hours. Participants will drive a driving simulator and perform in-vehicle tasks like tuning a radio or talking on a cell phone. If selected, you will be invited to come to a laboratory facility located on the Transportation Research Center Proving Grounds in East Liberty, Ohio. Participants will receive $42 per hour for time spent at the data collection facility, as well as mileage reimbursement for travel to and from the data collection site.

Please apply online at Research Study Contact Information Form.

<u>Contact Information Form, Page 1</u>

Transportation Research Center, Inc.
Driving Simulator Research Study
Applicant Contact Information Form

Approved by Sterling IRB
IRB ID 4342

U.S. Department of Transportation
National Highway Traffic Safety Administration

This study is being conducted by the Transportation Research Center, Inc., for the United States Department of Transportation's National Highway Traffic Safety Administration (NHTSA) to evaluate different tools that researchers use to measure distraction caused by in-vehicle technologies and portable devices including cell phones.

Participation involves one session of approximately 7-8 hours. Participants will drive a driving simulator and perform in-vehicle tasks like tuning a radio or talking on a cell phone. If selected, you will be invited to come to a laboratory facility located on the Transportation Research Center Proving Grounds in East Liberty, Ohio.

If selected, you will receive $42 per hour for time spent at the data collection facility, as well as mileage reimbursement for travel to and from the data collection site.

NHTSA and TRC will not release any personal identifying information you provide. All information gathered will be kept confidential.

<u>Participation Requirements:</u>
* Must be 18 - 70 years old
* Good general health
* Must have a valid U.S. driver's license with no restrictions other than corrective lenses
* Drive at least 3000 miles per year
* Must have experience using a cell phone while driving

If you satisfy these criteria and are available to participate in a daytime session of approximately 7-8 hours sometime during the next 2 months, please enter the following information:

First Name: **Middle Initial:** **Last Name:**

E-Mail Address: **5 Digit ZIP Code:** **Phone number (best number to reach you during the day):**

Gender: **Age (years):**

Female

Contact Information Form, Page 2

Gender: **Age (years):**

- ○ Female
- ○ Male

Have you participated in a previous driving simulator study?

- ○ Yes
- ○ No

How did you learn about our study?

- ☐ Craigslist Ad
- ☐ Newspaper Print Ad
- ☐ Newspaper Web Site Online Ad
- ☐ Friend or Relative
- ☐ Facebook Ad
- ☐ Other (Please specify.):

Do you have a valid U.S. driver's license?

○ Yes ○ No

How many miles do you drive per year?

- ○ Less than 3000
- ○ 3000 to 7000
- ○ 7000 to 10000
- ○ 10000 to 15000
- ○ 15000 to 20000
- ○ 20000 to 30000
- ○ 30000 to 40000
- ○ More than 40000

After you click the Submit button, you will receive an email confirming receipt of your information. Your information will be evaluated by our research staff, and you will be informed by email within a few days regarding the status of your application. If we still need additional study participants in your category, you will be sent another email with a link to a form requesting more information.

If you are not invited to participate in this study, may we keep your contact information for possible participation in future studies? Uncheck the box below if you want us to delete your information from our data base.

Screening Form

The following three pages are the Screening Form. If the person loads the Screening Form by clicking on the personalized link in the email that is sent to them (after reviewing the contact information details), the person's name, email address, phone number, and ZIP code are filled in with the information the person provided in the contact information form.

Several of the fields are "hidden" when the form is first loaded, that is, they are conditional and only are displayed when the person's responses take on certain values. For example, all of the text boxes referenced by "Other (Please specify.):" only show up when that box is checked. The question "May we use text messaging to help with scheduling?" is only displayed if the person answers "Yes" to the question "Do you regularly communicate using text messages?" The request for the person's cell phone number and carrier is only displayed if the person answers "Yes" to the question "May we use text messaging to help with scheduling?"

Transportation Research Center, Inc.
Driving Simulator Research Study
Applicant Screening Form

www.nhtsa.gov

Transportation Research Center Inc

U.S. Department of Transportation
National Highway Traffic Safety Administration

This study is being conducted by the Transportation Research Center, Inc., for the United States Department of Transportation's National Highway Traffic Safety Administration (NHTSA) to evaluate different tools that researchers use to measure distraction caused by in-vehicle technologies and portable devices including cell phones.

Participation involves one session of approximately 7-8 hours. Participants will drive a driving simulator and perform in-vehicle tasks like tuning a radio or talking on a cell phone. If selected, you will be invited to come to a laboratory facility located on the Transportation Research Center Proving Grounds in East Liberty, Ohio.

If selected, you will receive $42 per hour for time spent at the data collection facility, as well as mileage reimbursement for travel to and from the data collection site.

NHTSA and TRC will not release any personal identifying information you provide. All information gathered will be kept confidential and stored in a password protected database on a local computer. Responses to health related questions will not be kept - they are only being asked to determine your eligibility for participation.

Please enter the information requested below:

Personal Information

First Name: **Middle Initial:** **Last Name:**

E-Mail Address: **Phone number (best number to reach you during the day):**

Date of birth (Please select, year, month, and day from the dropdown lists):

Year: **Month:** **Day:**

Home Address:

Street Address (first line, required):

Street Address (Second Line, Optional):

City:

State: Ohio **5 digit ZIP Code:** **ZIP+4 (optional):**

The address you provide must be a physical address, not a P.O. box number. This is necessary so that we may calculate the mileage payment for you from your home to TRC, Inc.

Driving Experience

	Year	Make	Model
What kind of vehicle do you normally drive?			

How comfortable are you at multi-tasking while driving (e.g., eating, drinking, changing radio stations, talking on a cell phone, talking with passengers)?

○ 0 ○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7 ○ 8 ◉ 9 ○ 10

0: Not at all comfortable 10: Very much at ease

Do you use a cell phone while driving?

◉ Yes ○ No

Do you use a navigation system, computer, or any other similar devices in your car?

◉ Yes ○ No

Medical History and Present Health Condition

The next section contains questions about your health and physical condition. This information will not be kept and is only asked to determine your fitness for participation in the study.

Do you have any health problems that affect driving?

◉ Yes ○ No

Please describe:

Do you wear prescription eye glasses or contacts while driving?

◉ Yes ○ No

Do you require reading glasses to use a cell phone while driving?

○ Yes ◉ No

Are you able to drive without the use of assistive devices?

◉ Yes ○ No

Do you have high blood pressure that is not controlled by medicine?

○ Yes ◉ No

Do you have a history of seizures or epilepsy?

○ Yes ◉ No

Are you susceptible to motion sickness?

○ Yes ◉ No

Do you have difficulty hearing and understanding normal conversation?

○ Yes ◉ No

Do you have any inner ear, dizziness, vertigo, or balance problems?

○ Yes ◉ No

Do you have diabetes for which insulin is required?

○ Yes ◉ No

Have you ever had a concussion, brain injury, or other injury resulting in decreased motor control or cognitive ability?

◉ Yes ○ No

Please describe:

Are you taking any medications (over-the-counter or prescription) that may cause drowsiness or affect your driving ability?

○ Yes ◉ No

Do you currently have any medical condition that might affect your ability to concentrate while driving, such as Attention Deficit Hyperactivity Disorder (ADHD), depression, anxiety, or claustrophobia?

◉ Yes ○ No

Please describe:

Availability

Please indicate which days you are available to participate in a session of 7-8 hours. Check all that apply:

	Any Day	Monday	Tuesday	Wednesday	Thursday	Friday
Days Available:	☐	☐	☐	☐	☐	☐

Is there anything else we need to know about your availability, such as dates on which you absolutely CAN'T come, or the minimum amount of notice you need in order to be able to schedule an appointment?

◉ Yes ○ No, I'm fairly flexible.

Please specify:

Are you available on short notice to participate in our study? Could we call you on the same day to schedule if necessary?

◉ Yes ○ No

May we use email to help with scheduling?

◉ Yes ○ No

Do you regularly communicate using text messages?

◉ Yes ○ No

May we use text messaging to help with scheduling?

◉ Yes ○ No

What is your cell phone number? ☐ The number listed at the beginning of this form is my cell phone number.

Who is your cell phone carrier?

○ AT&T	○ Sprint	○ Verizon
○ Nextel	○ Qwest	○ Cincinnati Bell
○ Virgin Mobile	○ T-Mobile	○ TracFone
○ Cingular	○ BellSouth	○ Alltel
○ Boost Mobile	○ Consumer Cellular	○ Net10 Wireless
○ Revol Wireless	○ nTelos Wireless	◉ Other

Please specify:

The name of your cell phone carrier is needed in order to send text messages to your cell phone from a computer using SMS messaging, since the SMS Gateway to be used depends on the carrier.

How long would you like to be considered for this study?

After you click the Submit button, you will receive an email confirming receipt of your information. Your information will be evaluated by our research staff. The study is expected to last several weeks. We will notify you as soon as we can whether we will be able to include you in this driving study. However, it might take a couple of weeks to make that determination.

safercar.gov

Submit

Appendix B: Instruction Materials

Briefing Room Checklist:

Tasks	**Training Description, Orientation in Briefing Room**
Informed Consent	Subject reads and signs ICF.
Study Description	Read *Study Description Overview*.

Simulator Venue Training Protocol Checklist:

Tasks	**Training Description, Simulator Protocol / Venue**
DS Overview	Read *Simulator Orientation*.
Task Performance	Read *Task Performance Feedback Description* (give copy of table to subject to look at). As required, once minimum.
Secondary Task Overview Training	Read *Secondary Task Instructions Overview* before proceeding with first secondary task training and practice. Then, do training and practice for each secondary task. As required, once minimum.
DRT Training & Practice	Read *Detection Response Task (DRT) Training*. Practice each DRT (30s or more, make sure of good response). As required, once minimum.
Driving Task Training	Read *Simulator Driving Task Instructions*.
Simulator Familiarization	No lead vehicle. Use/Read *Simulator and Driving Task Familiarization Drive Information for Experimenters*. [Files needed: Fam.evt, Constant50.Om]
Driving Task Familiarization	Adding lead vehicle. Use/Read *Simulator and Driving Task Familiarization Drive Information for Experimenters*. [Files needed: FamCF.evt, Constant50.Om]
Task Performance Feedback	Provide performance feedback on familiarization drive. CF & lane keeping performance: ___________________________.
Main Trials	Proceed to main experimenter sheet for secondary task practice and main trial performance. Subject can take breaks if/when needed, as appropriate.

Static (DRT) Venue Training Protocol Checklist:

Tasks	**Training Description, Static (DRT) Protocol / Venue**
DS Overview	Read *Static Vehicle Orientation*.
Task Performance	Read *Task Performance Feedback Description* (give copy of table to subject to look at). As required, once minimum.
Secondary Task Overview Training	Read *Secondary Task Instructions Overview* before proceeding with first secondary task training and practice. Then, do training and practice for each secondary task. As required, once minimum.
DRT Training & Practice	Read *Detection Response Task (DRT) Training*. Practice each DRT (30s or more, make sure of good response). As required, once minimum.
Main Trials	Proceed to main experimenter sheet for secondary task practice and main trial performance. Subject can take breaks if/when needed, as appropriate.

STUDY DESCRIPTION OVERVIEW

Thanks for agreeing to participate. You will have a number of breaks including a longer break halfway through the experiment. Please feel free to ask questions at any time and let us know if you need a break.

The experiment will involve approximately 48 trials, half of which are practice and half are for testing. Each trial will last about 4 minutes. We will give you specific instructions before each trial. Please make sure that you don't start a trial if you are confused or don't know what we are asking you to do.

In each trial, you will perform an in-vehicle task and a detection response task in one of two test venues.

In-Vehicle Tasks:

The two in-vehicle tasks used today will be: (1) manually tuning an in-vehicle radio, and (2) performing a verbal task that is similar to hands-free cell phone conversation.

Detection Response Tasks:

Detection response tasks (DRTs) involve a timed sequence of stimuli, each requiring a button-press response. The stimuli will be either LEDs that illuminate periodically or a small vibrator placed on your skin. Three different DRTs will be used today. In the head-mounted DRT, an LED is attached to a light-weight device worn on your head. In the remote DRT, an LED is positioned at a remote location in front of the vehicle's steering wheel. The third DRT is called the tactile DRT; a small electrical vibrator will be temporarily attached to your shoulder using medical tape. Button press responses will be made using a micro-switch that is attached to your left index finger.

Test Venues:

In the "Driving Simulator" test venue, you will perform in-vehicle tasks and the DRTs while driving a simulated vehicle. In the "Static" test venue, you will perform in-vehicle tasks and the DRTs with no driving task.

Driving Simulator:

The driving simulator used in this study is a fixed-base simulator. It does not move. The simulator is connected to a 2010 Toyota Prius. While driving the simulator, you will sit in the driver's seat. You will control the simulator by moving the steering wheel and the gas and brake pedals of the study vehicle. The vehicle will have its engine turned off. A large screen in front of the vehicle will display a computer-generated image of the road on which you will be driving.

Do you have any questions so far?

Test Venue Orientation

SIMULATOR ORIENTATION

This vehicle is a Toyota Prius, which has been modified to collect driving performance data. You will be sitting in this vehicle to drive the simulator. Please get into the driver's seat and adjust the seat to your comfort. Make sure that you can reach the buttons on the center console and the task screen located to your right. The seat controls are under the front and on the lower left side of the seat. There is no need to adjust the mirrors as you will not be using them for this experiment. No shifting is required in this vehicle.

We have added sensors to the steering wheel, accelerator and brake pedals. These sensors allow us to run the driving simulator without having the vehicle turned on. Your control inputs are recorded by these sensors and input to the simulator to change the roadway image projected on the screen in front of you.

While driving in the simulator, remember, safe driving is the highest priority! You should do your best to keep your vehicle centered in the designated travel lane at all times and to maintain a constant following distance behind the lead vehicle. Car following and lane keeping performance are both measured as part of the primary task of driving. The car following task will be explained to you in a few minutes.

Do you have the seat adjusted the way you like it?

***

STATIC VEHICLE ORIENTATION

This vehicle is a Toyota Prius, which has been modified to collect test performance data. Please get into the driver's seat and adjust the seat to your comfort. The seat controls are under the front and on the lower left side of the seat. Make sure you can reach the buttons on the center console and the task screen located to your right, for we will be using the center console and task screen to perform some secondary tasks while seated in this vehicle. In this vehicle, you will not need the mirrors or vehicle controls.

Do you have the seat adjusted the way you like it?

TASK PERFORMANCE FEEDBACK DESCRIPTION

This table defines three levels of performance for the driving task components of the simulator test venue; as well as the detection response and secondary tasks that are used in both test venues. For the static test venue, only the detection response and secondary tasks are performed. All of these tasks will be explained later, but for now we will review the general performance criteria.

Task	**Good Performance**	**Acceptable Performance**	**Poor Performance**
Car Following	Maintains following distance consistently with minor deviations	Maintains following distance mostly with some noticeable deviations	Generally fails to maintain following distance
Lane Keeping	Maintains lane position consistently with minor deviations	Maintains lane position mostly with some noticeable deviations	Generally fails to maintain lane position
Detection Response Task (DRT)	Consistently attentive to DRT detection, detecting most stimuli	Moderate number of DRT stimuli not detected	Fails to detect significant number of DRT stimuli
Secondary Tasks	Performs secondary task continuously with minimal errors	Performs secondary task either intermittently or with moderate number of errors	Performs secondary task with considerable difficulty, slowly, and with moderate number of errors

Do you have any questions about the performance feedback?

SIMULATOR DRIVING TASK INSTRUCTIONS

Our simulator is a fixed-base driving simulator, meaning that it has no motion. The simulated driving environment will be a 4-lane roadway with a lead vehicle traveling in front of you.

When the roadway image first appears, your vehicle will be stopped and you should accelerate to 50 mph. After several seconds, a vehicle will appear ahead of you in your travel lane. We call this the "lead vehicle" because it is leading you in the car following task. Your task is to follow this vehicle, adjusting your speed as necessary to maintain a constant following distance. When the lead vehicle first appears, it will be 220 feet ahead of you. This is the desired following distance. You should take note of this distance when the vehicle first appears on the screen and try to maintain this following distance throughout the entire drive. The lead vehicle will maintain a constant speed of 50 mph throughout the drive.

Remember, safe driving is your highest priority! Both car following and lane keeping performance are measured as part of the primary task of driving. You should keep the vehicle in the center of the right lane and do your best to maintain a following distance of 220 feet behind the lead vehicle. If your following distance increases, it is OK to drive faster than 50 mph to catch up to the lead vehicle. If your following distance decreases, it is OK to drive slower than 50 mph to return to the specified following distance.

On each trial, you will drive approximately 3 miles. You should continue driving and performing the secondary task until the lead vehicle disappears, which signifies the end of the trial. Shortly thereafter, the simulator screen will shut off and go blank.

Do you have any questions or need a repeat of any instructions about the driving simulator or car following tasks before we practice?

SIMULATOR AND DRIVING TASK FAMILIARIZATION DRIVE
INFORMATION FOR EXPERIMENTERS

SIMULATOR FAMILIARIZATION DRIVE (lead vehicle is not present, no CF (Fam.evt))

When participant is ready to drive simulator:

> *"This drive is your 'test drive.' We want you to get a feel for driving in the simulator. The road will be straight except for one initial curve. There will be no other traffic or in-vehicle tasks. Remember to keep your hands and feet off the controls until the roadway image appears.*
>
> *When the roadway image appears, you may begin to press the accelerator and steer the vehicle. Speed up to about 50 mph and then slow down using the brake. Try making a lane change, then try keeping the vehicle centered in the travel lane for a while. Try maintaining a constant speed. Do whatever you need to become comfortable driving the simulator."*

When drive is over,

> *"Ok. Do you have any questions or do you want to practice this drive again?"*

DRIVING TASK FAMILIARIZATION DRIVE (car following, lead vehicle (FamCF.evt))

> *"In the next drive, we will add the car following task."*
>
> *"This drive will begin like the last one, but shortly after you get around the initial curve a lead vehicle will appear ahead of you in your travel lane. Make sure you are driving at approximately 50 mph when the lead vehicle appears because that is the initial speed of the lead vehicle. Remember to make note of the distance between your vehicle and the lead vehicle when it first appears, as this is the desired following distance that you should try to maintain throughout the drive."*
>
> *"The speed limit sign says 50 mph, but you can drive faster to catch up to the lead vehicle if you fall behind. In our scoring, your ability to maintain the designated following distance is our primary performance measure. You should also drive in the right lane and try to keep the vehicle centered in that lane at all times."*

After drive, provide performance feedback:
  Following distance and lane keeping performance
  Repeat driving task instructions as needed

Subjects can repeat this practice drive as needed, and should repeat if they have any difficulty, such as poor car following performance:

> *"Ok. Do you have any questions about the car following and lane keeping task or do you want to practice this drive again before we move on?"*

DETECTION RESPONSE TASK (DRT) TRAINING

While performing each secondary task, you will be asked to respond to a detection response task (DRT), which requires you to respond to a sequence of simple stimuli that will be presented to you one at a time. You will respond to a stimulus by pressing a micro-switch that will be attached to your left finger. The micro-switch is attached by wire to our data acquisition system. This equipment allows us to record the time at which each response is made. Today, we have three different detection response task types to test, one at a time.

The three detection response task types are: a head-mounted DRT, a tactile DRT, and a remote DRT. I will show you each of these devices and let you practice them momentarily, but first I will explain the stimulus and response method.

The stimuli are either red LED lights or a localized vibration, depending upon the task type. When you see or feel the stimulus, you should respond as quickly as possible by pressing the micro-switch attached to your finger. A stimulus will be presented every 3 to 5 seconds and will remain on until the button is pressed, or for about 1 second if no response is made. You will be scored based on your speed and accuracy in detecting the stimuli.

Now, I will show you each of the detection response task types and allow you to try them. First, please place the response button on your left index finger such that the button is comfortable and can be pressed while you are holding the steering wheel.

> [In the simulator venue] Please respond to a stimulus by pressing the button against the steering wheel, and use this method (pressing against the steering wheel) consistently throughout this test venue.
>
> [In the static test venue] You may press the button against either the steering wheel or against your thumb. I do ask that you choose a method (thumb or steering wheel) during this training and then use it consistently throughout this test venue.

[Exp: Make sure button and wire are positioned correctly, on left index or middle finger.]

Ok, here's the first one to try, the Remote DRT.

> Go ahead and try a few button presses in response to the stimuli. If you press the button quickly, a stimulus will shut off. If you do not respond quickly, it shuts off after 1 second.

And now let's try the next one, the Head-mounted DRT…
And, here's the final one, the Tactile DRT…

Do you have any questions about these detection response tasks? You will be given the opportunity to practice these DRTs again, in combination with the secondary tasks before the main trials.

SECONDARY TASK INSTRUCTIONS OVERVIEW

In the simulator, car following and lane keeping are considered the “Primary Task,” because safe vehicle control is your primary responsibility. In both test venues, you will also perform detection response tasks and “Secondary Tasks.” In the simulator, performing secondary tasks can interfere with car following and lane keeping, but it is important that you don’t let primary task performance deteriorate too much while performing a secondary task.

Instructions for performing secondary tasks will be presented auditorily so you don’t have to look for this information. For some tasks, the information will also be displayed on the computer screen located to the right of the center console, in case you forget. We call this the Task Screen. The Task Screen is a “touch screen,” which means that you will touch or press it when you complete each task.

In the simulator, the first secondary task will be presented shortly after the lead vehicle appears. In the static test venue, we will press a start button to initiate the secondary task. In both test venues, you will perform secondary tasks continuously over a trial that will last several minutes.

For some tasks, you will work continuously and will not hear additional instructions. For these tasks, information is presented auditorily throughout the trial and you will respond verbally. No information will be presented on the Task Screen.

Other secondary tasks have well-defined beginning and end points and require using manual controls and looking at in-vehicle displays. For these tasks, you will hear additional instructions and information. Such information will also be presented on the Task Screen.

Don’t worry if you make an error. We don’t expect perfect performance. If you make an error while performing a secondary task, please try to correct it before moving on. We will provide specific information about how to correct errors. It is important that you try to complete each task if possible.

Do you have any questions or need a repeat of any instructions before we move on to training for today’s first secondary task?

RADIO TUNING INSTRUCTIONS

In this task you will tune the radio to a designated frequency by using the tuning knob at the upper right corner and the buttons on the left side of the radio/navigation module. This vehicle has buttons on the steering wheel and a touch screen interface, but we ask that you do not use those features for this task. During each test venue, you will select several different radio frequencies, one at a time. You will be given the band (AM or FM) and the frequency each time.

In the simulator venue, the first frequency will be presented shortly after the lead vehicle appears. In the static test venue, we will begin the trial when you are ready. At the beginning of a trial, you will hear the first frequency followed by the word "BEGIN". At this point, you should work quickly and accurately to complete the task.

First, press the "AUDIO" button at the bottom of the column of buttons to the left of the vehicle's video screen. The audio display will then appear on the video screen.

Next, select the frequency band by pressing the AM or FM button located to left of the video screen. (Please use the buttons for this task, not the on-screen AM/FM tabs.) The current band will be displayed in the upper left of the screen and current frequency in the upper right. If you select the wrong band, press the correct button for the appropriate band. (After about 20 seconds of inactivity, the display will revert to a different screen. If this occurs, press the "AUDIO" button again to return to the audio screen.)

Use the tuning knob, located to the upper right of the screen, to adjust the frequency. When you have reached the specified frequency, say "DONE" aloud, and then press the "DONE" button on the Task Screen to complete the radio tuning task.

At this point, we want you to return the system to the original condition. To do this, you will press the "MAP" button to the right of the vehicle's video screen to take us out of radio mode. Then press the "NEXT" button on the Task Screen. Once you press the "NEXT" button, you should hear and see the next radio frequency. You will then perform the same sequence, starting with the "AUDIO" button. You will continue in this way until the trial is complete.

If you select the wrong band or frequency, try to fix it before moving on. If you notice an error but have already said "DONE", you do not need to try to fix it. You should continue on to the next step on the Task Screen.

Do you have any questions or need a repeat of any instructions before we practice this task?

[Load RadioTrain.tsv for stationary practice.]

N-BACK TASK INSTRUCTIONS

The “n-back” is an auditory memory task. In this task, you will hear a voice recording of a sequence of numbers presented one at a time with a couple seconds between each number. The numbers are separated into groups of 10, with separation between the groups consisting of a brief silence period followed by the word “Next’. This sequence continues for your entire driving or static task interval, in which the silent periods give you a momentary break from the task. Your task will be to remember the most recent numbers and say a specified number aloud after each presentation. The specified number will be either “0-back” or “1-back.”

An example of the sequence of numbers you will hear is presented in the left-most column in each table below. First you will hear “4” then “6” then “7” and so on. The responses that you should say aloud for the 0-back and 1-back conditions are presented in the right-most column of each table. Notice that in each condition, the sequence that you are to say aloud is the same as the original sequence. It is just delayed in the 1-back condition. Example Sequence:

0-Back Task	
What You Hear	What You Should Say
Next	nothing
4	4
6	6
7	7
3	3
1	1
2	2
9	9
5	5
8	8
0	0

1-Back Task	
What You Hear	What You Should Say
Next	nothing
4	nothing
6	4
7	6
3	7
1	3
2	1
9	2
5	9
8	5
0	8

Let’s look more closely at the 1-back condition: After the word “Next’, the first number you hear is “4.” Because there is no 1-back number at this point, you will not say anything. Next you will hear a “6.” At this point you will say “4” because it is the number that is one back from the current number 6. Next, you will hear “7” and you should say “6” because it is one back from the current number 7. With the exception of the first number occurring after the word “Next’, you will say a number aloud immediately after hearing each number in the 1-back condition. You should say the number quickly so that you don’t miss the next number, which will be presented within a couple seconds.

You will be given instructions before each drive about which version of the task you will perform (0-back or 1-back). Most people will make mistakes in the more difficult condition. If you become aware that you have made a mistake, it might help to clear your memory by not responding to the item (in 1-back) and effectively start over. Your performance score will be determined by the number of correct responses you make.

Do you have any questions before we practice this task? [Run 0Train.txt, then 1Train.txt for stationary practice.]

Appendix C: Participant Informed Consent Form

PARTICIPANT INFORMED CONSENT FORM

STUDY TITLE: Detection Response Task (DRT) Evaluation for Driver Distraction Measurement Application

STUDY INVESTIGATOR: Thomas A. Ranney, Ph.D.

STUDY SITE: Transportation Research Center, Inc.
10820 State Route 347
East Liberty, OH 43319

TELEPHONE: 800-[redacted]

SPONSOR: U.S. Department of Transportation
National Highway Traffic Safety Administration (NHTSA)

You are being asked to participate in a research study. Your participation in this research is strictly voluntary, meaning that you may or may not choose to take part. To decide whether or not you want to be part of this research, the risks and possible benefits of this study are described in this form so that you can make an informed decision. This process is known as informed consent. This consent form describes the purpose, procedures, possible benefits and risks of the study. This form also explains how your information will be used and who may see it.

The study investigator or study staff will answer any questions you may have about this form or about the study. Please read this document carefully and do not hesitate to ask anything about this information. This form may contain words that you do not understand. Please ask the study investigator or study staff to explain the words or information that you do not understand. After reading the consent form, if you would like to participate, you will be asked to sign this form. You will be offered a copy of the form to take home and keep for your records.

This research study is being conducted by the National Highway Traffic Safety Administration (NHTSA).

PURPOSE

The purpose of this study is to evaluate the different tools that researchers use to measure the level of distraction caused by "in-vehicle technologies." The latest in-vehicle technologies provide services such as Internet access, navigation information (maps and driving directions), as well as the ability to send and receive e-mails and text messages. Many in-vehicle systems allow such tasks to be performed with voice commands and auditory responses.

As new in-vehicle technologies are developed and marketed, there is a concern that these systems may interfere with driving. NHTSA is conducting this research study to determine the best way to collect data (information) on the use and impact of in-vehicle technologies while driving.

STUDY REQUIREMENTS

You are being asked to participate in this research study because:

1. You are 18 – 70 years of age,
2. You are an active driver with a valid, unrestricted U. S. driver's license (except for restrictions concerning corrective eyeglasses and contact lenses),
3. You drive at least 3,000 miles per year,
4. You are in good general health, and
5. You have experience using a wireless phone while driving.

NUMBER OF STUDY SITES AND STUDY PARTICIPANTS

This study will take place at one research site (Transportation Research Center, Inc.) and will include a minimum of 48 men and women.

STUDY PROCEDURES

Before participating in this research study, you will be asked to read this Participant Informed Consent Form in its entirety. After all of your questions have been answered, you will be asked to sign this form to show that you voluntarily consent to participate in this research study.

Your participation in this research study will consist of one session lasting approximately 7 – 8 hours. A member of the study staff will give you detailed instructions and will accompany you at all times during your participation in this research study. During the session you will complete approximately 48 test trials, each lasting approximately 3 – 4 minutes. In each trial, you will perform a combination of an in-vehicle task and a detection response task in one of two test venues. You will perform these task combinations while you are sitting in the driver's seat of a stationary vehicle. Details are presented in the following sections.

In-Vehicle Tasks:

The two in-vehicle tasks used in this study will consist of manually tuning a radio using an in-vehicle system and performing a verbal digit recall task that involves listening and speaking and is similar to a hands-free cell phone conversation.

Detection Response Tasks:

Detection response tasks (DRT) are used to measure the amount of distraction associated with an in-vehicle task. A DRT involves a timed sequence of artificial stimuli, each requiring a button-press response. Three variations of DRT will be used in this experiment. Two variations will use simple visual stimuli (light-emitting diodes or LEDs). One visual task variation will use an LED attached to a lightweight device worn on your head. The second visual task variation will use an LED that is positioned at a remote location in front of the vehicle's steering wheel. The third DRT variation will use a tactile stimulus; a small electrical vibrator will be temporarily attached to your shoulder using medical tape. Button press responses will be made using a micro-switch that is attached to your left index finger. Trials will never use more than one DRT variation.

Test Venues:

In the “Driving Simulator” test venue, you will perform in-vehicle tasks and the specified DRT while driving a simulated vehicle. In the “Static” test venue, you will perform in-vehicle tasks and the DRT with no driving task. All testing in one venue will be completed before moving to the second venue. You will have an opportunity to rest and eat between the two parts of the testing.

Driving Simulator

The driving simulator used in this study is a fixed-base simulator. A fixed-based simulator is a machine that imitates the conditions of driving in real life, but does not move. The simulator will be connected to a 2010 Toyota Prius. While driving the simulator, you will sit in the driver’s seat of the study vehicle. You will control the simulator by moving the steering wheel and the gas and brake pedals of the study vehicle. The vehicle will have its engine turned off. The vehicle used with the driving simulator is equipped with sensors to collect information on your steering, braking and gas pedal usage. The sensors are located so that they will not affect your driving. The information collected by these sensors is recorded so that it can be analyzed at a later time. A large screen in front of the vehicle will display a computer-generated image of the virtual road on which you will be driving.

While operating the simulator, you will be asked to perform specific driving tasks. These tasks will involve activities such as following a car at a specified distance and keeping the vehicle within the specified travel lane.

Summary of Study Procedures:

The following procedures will take place at your session:

- After signing this consent form, you will be provided instructions and training on driving the simulator, DRT, and performing the in-vehicle tasks. You will also be given the opportunity to practice each of these before performing test trials.
- You will then complete 2 sets of trials, including approximately 24 driving simulator trials each lasting 3.5 minutes and approximately 24 trials in the static test venue, each lasting approximately 3 minutes.
- After completing both driving simulator and static trials, the session will end and your participation in this research study will be complete.

NEW INFORMATION

We do not anticipate that any changes to procedures will take place during this study. However, any new information developed during the course of the research that may affect your willingness to participate will be provided to you.

RISKS of STUDY PARTICIPATION

Most people enjoy driving in the simulator and do not experience any discomfort. However, a small number of participants experience symptoms of discomfort associated with simulator disorientation. Previous studies with similar driving intensities and simulator setups have produced mild to moderate disorientation effects such as slight uneasiness, warmth, or eyestrain for a small number of participants. These effects typically last for only a short time, usually 10 – 15 minutes, after leaving the simulator. If you ask to stop driving as a result of discomfort, you will be allowed to stop at once. You will be asked to sit and rest before leaving. You will also be given the opportunity to consume a beverage and/or a snack. There is no evidence that driving ability is hampered in any way; therefore, if you show minimal or no signs of discomfort, you should be able to drive home. If you experience anything other than slight effects, transportation will be arranged through other means. This outcome is considered unlikely since studies in similar devices have shown only mild effects in recent investigations and evidence shows that symptoms decrease rapidly after simulator exposure is complete.

Participants will be asked to wear Detection Response Task equipment while driving. The head-mounted version requires wearing the suspension of a hard hat to which a flexible arm and single LED have been attached. It has no associated risk. The tactile version requires attaching a small electrical vibrator to the participant's shoulder with medical adhesive tape. The vibrator will be activated periodically for durations of one second. The level of vibration will be set to be noticeable but not uncomfortable. The associated risk is expected to be no more than mild discomfort for a small percentage of participants.

There are no known physical or psychological risks associated with participation in this study beyond these.

BENEFITS of STUDY PARTICIPATION

This research study will provide data on driver behavior and in-vehicle task performance that will be used by researchers to provide a scientific basis for developing recommendations or standards for performing in-vehicle tasks while driving. Your participation in this study will provide data that may help develop these recommendations or standards.

You are not expected to receive direct benefit from your participation in this research study.

ALTERNATIVES

This study is for research purposes only. Your alternative is to not participate.

CONDITIONS OF PARTICPATION, WITHDRAWAL, AND TERMINATION

Participation in this research is voluntary. By agreeing to participate, you agree to operate the research equipment in accordance with all instructions provided by the study staff. If you fail to follow instructions, or if you behave in a dangerous manner, you may be terminated from the study. You may withdraw your consent and discontinue participation in the study at any time without penalty.

COSTS TO YOU

Other that the time you contribute, there will be no costs to you.

COMPENSATION

You will receive $42.00 per hour for the time you spend at the data collection facility. You will receive mileage reimbursement for travel to and from the data collection site.

If you voluntarily withdraw or are terminated from this study, you will be compensated for the number of hours that you participated in the study.

USE OF INFORMATION COLLECTED

In the course of this study, the following data will be collected:

- Engineering data (such as the information recorded by the study vehicle sensors)
- Video/audio data (such as the information recorded by the video cameras)

Information NHTSA may release:

The engineering data collected and recorded in this study will include performance scores based on the data. This data will be analyzed along with data gathered from other participants. NHTSA may publicly release this data in final reports or other publication or media for scientific, educational, research or outreach purposes.

The video/audio data recorded in this study includes your video-recorded likeness and all in-vehicle audio (including your voice). The video/audio data may include information regarding your driving performance. Video and in-vehicle audio will be used to examine your driving performance and other task performance while driving. NHTSA may publicly release video image data (in continuous video or still formats) and associated audio data, either separately or in association with the appropriate engineering data for scientific, educational, research or outreach purposes.

Information NHTSA may not release:

Any release of engineering data or video/audio data shall not include release of your name. However, in the event of a court action, NHTSA may not be able to prevent release of your name or other personal identifying information. NHTSA will not release any information collected regarding your health and driving record.

QUESTIONS

Any questions you have about the study can be answered by Thomas Ranney, Ph.D., or the study staff by calling [redacted].

If you have questions regarding your rights as a research participant, or if you have questions, concerns, complaints about the research, would like information, or would like to offer input, you may contact the Sterling Institutional Review Board Regulatory Department, 6300 Powers Ferry Road, Suite 600-351, Atlanta, Georgia 30339 (mailing address) at telephone number [redacted] (toll free).

INFORMED CONSENT

By signing the informed consent statement contained in this document, you agree that your participation is voluntary and that the terms of this agreement have been explained to you. Also, by signing the informed consent statement, you agree to operate the study equipment in accordance with all instructions provided by the study staff. You may withdraw your consent and discontinue participation in the study at any time without penalty.

NHTSA will retain a signed copy of this Informed Consent form. A copy of this form will also be offered to you.

INFORMED CONSENT STATEMENT

I certify that:

- I have a valid, U. S. driver's license.
- All personal and vehicle information, as well as information regarding my normal daily driving habits provided by me to NHTSA, and/or Transportation Research Center Inc. employees associated with this study during the pre-participation screening and the introductory briefing was true and accurate to the best of my knowledge.
- I have been informed about the study in which I am about to participate.
- I have been told how much time and compensation are involved.
- I have been told that the purpose of this study is to evaluate the tools that researchers use to measure driving and in-vehicle task performance.
- I agree to operate the research equipment in accordance with all instructions provided to me by the study staff.

I have been told that:

- Part of the study will be conducted in a fixed-base driving simulator and that the risk of discomfort associated with simulator disorientation is minimal.
- For scientific, educational, research, or outreach purposes, video images of my driving, which will contain views of my face and accompanying audio data, may be used or disclosed by NHTSA, but my name and any health data or driving record information will not be used or disclosed by NHTSA.
- My participation is voluntary, and I may refuse to participate or withdraw my consent and stop taking part at any time without penalty or loss of benefits to which I may be entitled.
- I have the right to ask questions at any time and that I may contact the study investigator, Thomas Ranney, Ph.D., or the study staff at [redacted] for information about the study and my rights.

I have been given adequate time to read this informed consent form. I hereby consent to take part in this research study. I do not give up any of my legal rights by participating in this study.

You will receive a signed copy of this Participant Informed Consent Form, which has 9 pages.

I, ____, voluntarily consent to participate.
(*Printed Name of Participant*)

____
Signature of Participant Date

INFORMATION DISCLOSURE

By signing the information disclosure statement contained in this document, you agree that the National Highway Traffic Safety Administration (NHTSA) and its authorized contractors and agents will have the right to use the NHTSA engineering data and the NHTSA video and audio data for scientific, educational, research, or outreach purposes, including dissemination or publication of your likeness in video or still photo format, but that neither NHTSA nor its authorized contractors or agents shall release your name; and you have been told that, in the event of court action, NHTSA may not be able to prevent release of your name or other personal identifying information. NHTSA will not release any information collected regarding your health and driving record, either by questionnaire or medical examination. Your permission to disclose this information will not expire on a specific date.

Information Disclosure Statement

I, ____, grant permission to
(*Printed Name of Participant*)

the National Highway Traffic Safety Administration (NHTSA) to use, publish, or otherwise disseminate NHTSA engineering data and NHTSA video image data, as defined in the Participant Informed Consent Form (including continuous video and still photo formats derived from the video recording), and associated with the appropriate engineering data for scientific, educational, research or outreach purposes. I have been told that such use may involve widespread distribution to the public and may involve dissemination of my likeness in video or still photo formats, but will not result in release of my name or other identifying personal information by NHTSA or its authorized contractors or agents. I have been told that my permission to disclose this information will not expire on a specific date.

____
Signature of Participant Date

Appendix D: Simulator Sickness Questionnaire

Participant Number: ___

Directions:

Circle one option for each symptom to indicate whether that symptom applies to you right now.

1. General Discomfort...............None..............Slight..............Moderate ....Severe
2. Fatigue ....None..............Slight..............Moderate ....Severe
3. Headache ....None..............Slight..............Moderate ....Severe
4. Eye Strain ....None..............Slight..............Moderate ....Severe
5. Difficulty Focusing ....None..............Slight..............Moderate ....Severe
6. Salivation Increased ....None..............Slight..............Moderate ....Severe
7. Sweating ....None..............Slight..............Moderate ....Severe
8. Nausea ....None..............Slight..............Moderate ....Severe
9. Difficulty Concentrating ....None..............Slight..............Moderate ....Severe
10. "Fullness of the Head" ....None..............Slight..............Moderate ....Severe
11. Blurred Vision ....None..............Slight..............Moderate ....Severe
12. Dizziness with Eyes Open ......None..............Slight..............Moderate ....Severe
13. Dizziness with Eyes Closed ....None..............Slight..............Moderate ....Severe
14. *Vertigo ....None..............Slight..............Moderate ....Severe
15. **Stomach Awareness ....None..............Slight..............Moderate ....Severe
16. Burping.................................No..................Yes.................If yes, no. of times ______
17. ....Vomiting.........No..................Yes If yes, no. of times ______
18. Other ______________________________________

* Vertigo is experienced as loss of orientation with respect to vertical upright.

** Stomach awareness is usually used to indicate a feeling of discomfort which is just short of nausea.

DOT HS 812 077
November 2014

U.S. Department
of Transportation

National Highway
Traffic Safety
Administration

NHTSA
www.nhtsa.gov

10999-110514-v2