Self-Regulation Minimizes Crash Risk from Attentional Effects of Cognitive Load during Auditory-Vocal Tasks

Richard Young
Wayne State Univ.

ABSTRACT

This study reanalyzes the data from a recent experimental report from the University of Utah investigating the effect on driving performance of auditory-vocal secondary tasks (such as cell phone and passenger conversations, speech-to-text, and a complex artificial cognitive task). The current objective is to estimate the relative risk of crashes associated with such auditory-vocal tasks. Contrary to the Utah study's assumption of an increase in crash risk from the attentional effects of cognitive load, a deeper analysis of the Utah data shows that driver self-regulation provides an effective countermeasure that offsets possible increases in crash risk. For example, drivers self-regulated their following distances to compensate for the slight increases in brake response time while performing auditory-vocal tasks. This new finding is supported by naturalistic driving data showing that cell phone conversation does not increase crash risk above that of normal baseline driving. The Utah data are next compared to those from a larger study that included visual-manual as well as auditory-vocal tasks. The Utah auditory-vocal tasks had negligible effects on response time compared to visual-manual tasks with socially acceptable crash risk, such as manual radio tuning. In conclusion, auditory-vocal tasks such as those in the Utah study are not expected to increase crash risk or impair real-world operation of an automobile, compared to normal baseline driving without performing such tasks.


OUTLINE

Part 1 reviews a recent experimental study of auditory-vocal tasks conducted by researchers at the University of Utah [1] (hereafter called the “Utah” study). Its “Cognitive Distraction Index” (hereafter called the “Index”) for auditory-vocal tasks is reviewed. In Part 1, the original data for key variables making up this Index are re-analyzed. The Index scores are compared to the real-world relative crash risk for cellular conversations in naturalistic driving. Part 1 examines whether the Utah data support its assumption that auditory-vocal tasks increase crash risk, or, support the alternate view that driver self-regulation maintains crash risk at normal driving levels during auditory-vocal tasks. Part 2 compares key data for the Utah auditory-vocal tasks to the Crash Avoidance Metrics Partnership Driver Workload Metrics (CAMP-DWM) data [2] for auditory-vocal and visual-manual tasks.

PART 1. THE UTAH STUDY

1.1 Introduction

1.1.1 Utah Cognitive Index

The Utah study used a large number of variables and test conditions to assess the attentional effects of cognitive load from auditory-vocal tasks on driving. It tested 7 auditory-vocal task conditions and 1 “single” (no secondary task) condition in 3 experimental venues: standalone with no driving, in a simulator while following a lead vehicle, and in an instrumented vehicle in on-road city traffic at about 25 mph. It gathered 31 dependent variables, condensed into 13 final variables by averaging the scores for the same variable collected in different venues. From this condensed set of 13 variables, the Utah investigators created the Index. Subjective workload scales constituted 6 of the final 13 variables and so were heavily weighted in the Index. The 13 variables were “standardized” to a common scale, and summed to create a single combined score, which was transformed into a 5-point scale by anchoring the low end at “Category 1” with the composite score for the “Single” baseline condition with no secondary task. The high end was anchored at “Category 5” with the composite score for a complex artificial task called “OSPAN” in which subjects attempted to recall single syllable words in serial order while solving mathematical problems [1, p. 13]. Figure 1 re-plots the final scores of the 8 task conditions as given in [1, Figure 28].
1.1.2 Crash Risk Assumption in the Utah Study

The Utah study [1, p. 30] assumes that its Index relates to crash risk, “It is reasonable to assume that there would be a monotonic relationship between cognitive distraction and crash risk.” (Although not explicitly stated, it is obvious that the Utah report assumes a monotonic increase, not a monotonic decrease, in crash risk with increased cognitive distraction.)

The Utah study placed its three conversation tasks - talking to a passenger, talking on a hand-held phone, and talking on a hands-free phone (Figure 1, middle bars) - into a “moderate” Category 2 cognitive demand level. The study concludes that conversation tasks, “…are significant impairments to driving that stem from the diversion of attention from the task of operating a motor vehicle.” [1, p. 30]

Another task called “speech-to-text” involved a simulated voice interaction with a machine rather than a human. The Utah study concludes that this task (with a score of 3.06; second-to-last bar in Figure 1) was even more cognitively demanding than conversation, and so placed it in a higher “Category 3” Index level. The report states [1, p. 29]:

Given the current trends toward more voice commands in the vehicle, this Category 3 level of cognitive distraction is troubling. The assumption that if the eyes were on the road and the hands were on the steering wheel then voice-based interactions would be safe appears to be unwarranted.

The study concludes [1, p. 30] that adoption of voice-based speech-to-text and text-to-speech systems in the vehicle, “…may have unintended consequences that adversely affect traffic safety.

The report identified what it considers some potential driving “costs” incurred with doing the auditory-vocal tasks it tested, and attempted to reflect these “costs” in its Index (Figure 1). It refers 13 times to these costs as “impairments” to the driver, which in some cases “may rise to the level associated with drunk driving” [1, p. 29]. The use of a 5-category scale deliberately modeled on the Saffir-Simpson hurricane scale implies that safety concerns approach catastrophic levels at the upper end of the scale. The average reader is left with the impression that auditory-vocal tasks performed while driving increase crash risk by enormous amounts, even to the level associated with drunk driving [1, p. 29].

However, there were no reported crashes on the road or in the simulator in the Utah study, and no mention of the number or percentage of experimenter interventions to alert the drivers to the presence of a roadway hazard on the road, belying the implication that a Category 5 distraction has enormous absolute crash risk. There was no evidence presented comparing the Index to known major impairments, such as drowsiness, drugs, or alcohol. No visual-manual tasks were tested for comparison to the auditory-vocal tasks. Indeed, prior experimental studies [2, 3] found that speech interfaces were less distracting than visual-manual interfaces. Without such a context, the upper bound of the Utah scale is unknowable. A higher upper bound would dramatically affect the relative significance of what is being expressed by the categories on the Utah scale. Hence, the validity of the Utah study’s assumptions, data, results, and conclusions about its Index bear scrutiny.

1.1.3 Driver Self-Regulation

An important fact not recognized in the Utah study [1] is that drivers routinely self-regulate their behavior during real-world driving. Drivers can, and do, “actively adjust their driving behavior in response to change or competing task demands to maintain an adequate level of safe driving” [3, 38]. Such self-regulation can occur from drivers choosing not to do a secondary task while driving, or adjusting their driving behavior while performing the task to keep their safety margins equal to or better than when not doing the task. Self-regulation is typically not observed in experimental studies, because drivers are not given a choice about performing a task (i.e., they must perform the tasks to meet the requirements of the experiment).

Many real-world non-experimental studies have observed such self-regulatory behavior in everyday driving, and proposed that drivers self-regulate their driving to compensate for the attentional effects of cognitive load and thereby reduce crash risk. For example, the compensation hypothesis was proposed by Young and Schreiner [4] to explain their finding that using the OnStar embedded device to make hands-free personal calls does not elevate the rate of crashes severe enough to deploy an airbag [4]. The compensation hypothesis states that drivers talking on a cell phone during natural driving tend to change their driving behavior such that there is no net safety decrement from the slightly increased response time (RT) to events during such conversations that is predicted from experimental studies [5, 6]. Indeed, there is previously unrecognized evidence of such self-regulation in the Utah study data, as shown in the current study.
1.2 Method

The means and standard errors for the “response time” and “following distance” variables were requested and received from the lead author of the Utah report [7]. These original data, rather than the “standard” scores in Appendix A of the Utah report [1], were used for the current in-depth analysis, because the “standard” scores were actually not standard scores as claimed in the Utah report. Across the 8 task conditions, each variable in its Appendix A has a mean of zero, but the standard deviations are not 1, as required for a standard score. The current study therefore used only the original data to investigate key variables in the Index.

1.3 Results

1.3.1 Brake RT to Brake Light Onset

Figure 2 plots the subject driver mean brake response time (RT) to the pace vehicle brake light onset in the Utah simulator experiment, reproducing Figure 23 in the original report [1]. Note that the brake RT in Figure 2 varies in the same manner as the Index scores in Figure 1 ($r = 0.95, p = 0.0004$). That is, the higher the brake RT, the higher the Index. This result is not surprising, since brake RT was one of the 13 variables used in the Utah report to create its Index. The largest RT was for the OSPAN task (1092 ms), which is a statistically significant increase of 171 ms, compared to the baseline “Single” condition (921 ms) ($t = 4.36, p = 0.00005$, independent two-sample t-test, equal sample sizes, equal variance). The increases for the three conversation tasks in the center of Figure 2 were 26 to 56 ms compared to the Single baseline (not statistically significant). These RT increases are near the low end of the 50-200 ms range (about the time of an eye blink) found in prior experimental venues (brain imaging, laboratory, simulator, test track, and open road), for the effect of conversation tasks on visual event RTs [2, 5, 6].

1.3.2 Following Distance to Pace Vehicle

Figure 3 plots the following distances in the Utah simulator experiment for the 8 task conditions. Note that the following distance also varies in a manner similar to the Index score in Figure 1 ($r = 0.92, p = 0.0014$). That is, the higher the Index, the greater the following distance. This result is again not surprising, since following distance was another of the 13 key variables on which the Index was based.

1.3.3 Increased Following Distance Compensates for Increased Brake RT

Previous studies [19, 20, 21, 22, 95], including those of the Utah investigators [11, 12, 13, 14, 15, 16, 17, 18], have interpreted increased following distance as driver self-regulation, providing an increased safety margin while performing auditory-vocal tasks. Should the lead vehicle suddenly apply its brakes and decelerate rapidly, increased following distance gives subjects more time to brake, thereby avoiding increased rear-end crash risk. The brake RT and following distance are well correlated ($r = 0.96, n = 8, p = 0.0002$), as is self-evident by comparing Figures 2 and 3.

Do these increases in following distance compensate for the increases in brake RT? The Utah study claims they do not: “Brake RT increased as a function of condition over and above any compensatory effects associated with following distance” [1, p. 18]. The basis for this conclusion was a “subsidiary linear mixed model analysis that held following distance constant,” but details were not provided of this model, its results, or why holding following distance constant would indicate it was not a compensatory behavior. Figure 3 shows that following distance was actually not constant in the Utah data, raising further questions of why the Utah model would hold it constant. A closer examination of following distance and its potential compensatory effects is therefore warranted.
One way to test if the observed increases in following distance (meters in Figure 3) were sufficient to compensate for the observed increases in brake RT (milliseconds in Figure 2) is to place the following distance on the same time scale as brake RT. This conversion is easy if the speed of the vehicle were known, because distance can then be transformed into time. However, the Utah investigators did not report the mean speed of the subject vehicles during the approximately 10-min task trials in the simulator. Therefore, transformation of the Utah study following distance to a time scale cannot be done with available information. (Reasonable speed assumptions for the Utah study give rise to similar results as the current study [8].)

In lieu of speed data, the compensation question can still be directly assessed by calculating the percentage changes in brake RT and following distance relative to the baseline “Single” condition. Speed affects only the absolute time headway; the relative time headway equals the relative distance headway, regardless of the average speed. For example, consider the following equations showing following distance D, speed S, and the headway time T for two tasks:

\[ D_1 = S_1 \times T_1 \]
\[ D_2 = S_2 \times T_2 \]

Taking the ratio of Equations 1 gives:

\[ \frac{D_1}{D_2} = \frac{S_1 \times T_1}{S_2 \times T_2} \]

(1)

The pace car had the same average speed for every task condition [9]. Otherwise, the driver performance metrics would have been confounded with speed in the experimental design, invalidating the Utah results. In addition, subjects would have tried to match their speed to the pace car as best they could in order to keep a fixed gap to the pace car as instructed. Hence, \( S_1 = S_2 \) and speed cancels out of Equation 2, leaving the ratio of the following distances equal to the ratio of the headway times:

\[ \frac{D_1}{D_2} = \frac{T_1}{T_2} \]

(2)

Equation 3 indicates that the ratio of the following distances D is equal to the ratio of the headway times T for the Utah simulator experiment. It follows that if the following distance increased proportionately to the brake RT for a given task (relative to the Single condition with no secondary task), then the subjects must have compensated for their increased brake RT by increasing their headway time (which was proportional to their following distance).

To test this simple deduction, the data in Figures 2 and 3 were each converted into percentages relative to their respective Single baseline. Figure 4 plots the percentage change in following distance (blue bars) vs. that for brake RT (red bars) on the same y-axis percentage scale for comparison.

For all 7 auditory-vocal tasks tested, Figure 4 shows that the following distance changed in a proportional manner (relative to the “Single” baseline), as did the brake RT. Not one task had a percentage increase in RT that was significantly larger than the percentage increase in following distance. Hence, under the conditions of the Utah simulator study, subjects compensated for the increases in brake RTs with proportionate increases in their following distances for all tasks in the study.

For example, Figure 2 shows that the RT increased from 921 ms for the “Single” control condition, to 1092 ms for the most difficult OPSAN task condition, a 171 ms or 18.6% increase (rightmost red bar in Figure 4). Figure 3 shows that the following distance increased from 23.3 m for the “Single” control condition, to 27.8 m for the most difficult OPSAN task condition, a 4.5 m or 19.3% increase (rightmost blue bar in Figure 4). Because these percentage increases were about the same, the increased following distance (and corresponding time headway) fully compensated for the increased RT.

Note that the subjects increased their following distances despite being instructed and trained, “to maintain a two-second following distance [sic] behind the pace car” [1, p. 21]. Evidently, drivers’ self-regulatory behavior was so strong that it even overcame the strong influence of “obedience to authority” (the instruction and training to maintain a constant 2-s following time for every task condition).
It is unknown if the increase in RTs from secondary auditory-vocal tasks in simulated driving would be large enough in real driving to affect driving safety, even in the absence of compensation from self-regulation. What Figure 4 demonstrates, however, is that the increased brake RTs in the Utah simulator study were fully offset by the increased following distances during those tasks, as evidenced by the Utah study's own data. Hence, the Utah study's data does not support its conclusion that the cognitive loads of auditory-vocal tasks cause "significant impairments to driving." In fact, the Utah data in Figure 4 shows no impairment to driving for any of the dual task conditions, relative to the Single task condition. Drivers maintained an adequate safety margin in all task conditions, due to self-regulation.

1.3.4 Estimating Crash Risk from Key Utah Index Variables

1.3.4.1 Brake Response Time and Crash Risk

The Utah study assumes [1, p. 30] that an increased Index score (of which brake RT is a key part) increases crash risk. That argument assumes the absence of compensatory adjustments from other factors. However, the specific relationship between crash risk and RT is not simple, and can be complicated by environmental conditions, road geometry, and many other factors [10]. For example, the brake RT of a following vehicle to a lead vehicle brake light onset is estimated from crash investigation data to be about 639 ms shorter when a driver is behind a lead vehicle in a curve vs. a straight road, despite the fact that crash risk is higher on curves vs. straight roads [10, Table 3]. That is, shorter brake RTs are associated with higher crash risk in curves, exactly the opposite of the Utah study monotonic increase assumption. This result is likely due to increased vigilance by drivers while negotiating curves, thus reducing their RTs, yet the crash risk is higher because of the more complicated road geometry.

1.3.4.2 Following Distance and Crash Risk

It is noteworthy that the Utah subjects increased their following distance despite the fact that they were in an experimental setting and had been trained and instructed beforehand to maintain a constant 2-s headway. The analysis in Figure 4 shows that the following distances of the drivers in the Utah study were sufficient to compensate for the increased RT for all tasks in the Utah study, even the most difficult “Category 5” OSPAN task. Increased time and/or distance headways during cell phone conversation or other auditory-vocal tasks is also evident in prior simulator studies by members of the Utah group [13, 14, 15, 16, 17, 18] and others [19, 20, 21, 22, 23], with only one known exception [23].

In naturalistic driving studies (NDS), drivers have no authority figure instructing them to maintain a constant headway. Indeed, drivers in NDS increase their time headway when conversing on three phone types: hand-held, portable hands-free, and integrated hands-free cell phone [24, Figure 18]. Fitch et al. [24, Table 94] reported that the mean time headway difference score for conversation vs. no conversation for each of the three phone types individually had p-values approaching significance at 0.095, and claimed there was no statistically significant effect of conversation on time headway. However, the current study pooled the p-values across the three cell phone types using the meta-analysis program in [25, and, contrary to Fitch et al.'s conclusion [17], found that there was a statistically significant effect of conversation at increasing time headways (p = 0.029). Thus, the real-world NDS data are consistent with the result of the current re-analysis of the Utah simulator data - drivers increase their following distances during cell phone conversation.

Do such increased headways reduce or increase crash risk? Some might argue that increased headways are caused by drivers not being able to perform the primary driving task properly under conditions of cognitive load, and therefore indicate "impairment" of the driver, which presumably increases crash risk. However, increased following distance is known to reduce real-world crash risk and traffic violations. For example, Evans and Wasielewski [26] measured headways of individual vehicles in real-world observational studies and then used each license plate number to look up the record of each vehicle's crash and ticket involvement, and found that increased headways were associated with fewer crashes and traffic violations. Hence, increased following distance (i.e., the reverse of tailgating) is a safety benefit in real driving. It is therefore invalid and misleading for the Utah report to include following distance in an Index that is assumed to correlate with increased crash risk, when in fact increases in following distance reduce crash risk.

1.3.4.3 Reduced Head Turns at Intersections

Locations with stop signs, stoplights, and/or crosswalks at intersections along the 25 mph speed limit city route of the Utah open road experiment were identified a priori by subjective judgment of the Utah investigators as posing a “potential” hazard. The Utah study calculated the probability of a left-right head turn by the subject driver at an intersection (presumably to fully scan for crossing traffic or pedestrians at those locations). (If the subject driver turned to look in just one direction but not the other, the Utah study marked it against the subject.) This head-turn probability was estimated from the video recordings in post-analysis and entered as a variable into the Index. If the probability (that the driver turned his/her head in both directions to scan the intersection) decreased during a task, the Utah study took that as indicating an effect of cognitive distraction. During the Passenger, Hand-Held, and Hands-Free conversation tasks, the percentage of scanning both directions at the intersection was about 66%. For the Single (baseline) condition it was about 74%, and for the most difficult OSPAN task it was about 61% [1, Figure 26]. This slight reduction in left-right scanning is partially consistent with the simulator results of Harbluk et al. [27, p. 272], who found that when approaching and driving through intersections, drivers reduced their scanning of intersection areas to the right.
Likewise, Gruzdaitis et al. [28] found a reduction in scans to the right at intersections in an urban on-road instrumented vehicle study with 15 younger subjects in Helsinki, Finland.

However, Fitch et al. [24, p. xxvii] in an NDS study of cell phone effects of driving performance state: “The likelihood of properly scanning an intersection was not found to differ when using a cell phone.” They found that drivers properly scanned the intersection during 38 percent of the baseline periods, 36 percent of the grouped visual-manual cell phone subtasks, and 35 percent of the grouped talking/listening cell phone subtasks. These small differences in scanning behavior are inconsistent with the Utah [1], Harbluk et al. [27], and Gruzdaitis et al. [28] results. In addition, the percentages of head turns (35-38%) in the Fitch et al. NDS study [24] are about half the percentages of head turns in the Utah on-road experimental study (61%-74%), for all task conditions. That is, the slightly reduced Utah head-turn probability at intersections when performing auditory-vocal tasks was still far higher than that observed during naturalistic baseline driving [24]. It is therefore difficult to argue that the observed percentage of head turns was indicative of increased crash risk compared to naturalistic driving. It is more plausible that the Utah study drivers were constantly aware of being observed by the experimenters in the car with them, and so were motivated to perform in the manner that they thought the experimenters expected them to perform, giving rise to far more head turns than would ordinarily occur in non-experimental real-world driving.

There are many other possible explanations for the slight 5-13% decrease in left-right head turns at intersections:

1. Drivers felt duty-bound to focus their gaze more on the pace vehicle than to scan for potential hazards that were never identified, in order to follow their explicit instructions and training to maintain a constant distance to the pace vehicle while performing the tasks.
2. Once the lead vehicle accelerated after stopping or slowing down at the intersection, subject drivers took that as indicating it was safe for them to proceed through the intersection, without the need to scan the intersection.
3. No actual hazard may have been present at any of the pre-defined locations during the experiment.
4. Even if an actual hazard were present, the Utah report does not mention if there were more left-right head turns during a secondary task than during no secondary task.
5. The Utah report does not mention whether there were more actual hazards during task conditions than during the “Single” baseline condition, or whether different task conditions had different numbers of actual hazards.
6. No instructions were given to the drivers regarding the need to search for hazards, and drivers were scored only on the “potential” hazards pre-identified by the experimenters, but never revealed to the subjects.
7. The current study used Google Earth to inspect the actual route used for the Utah on-road study [1, Appendix D] from a ground-level view, revealing that almost all intersections were free from physical obstruction of sight lines. Thus, when subject drivers were approaching an intersection, they could have seen from a distance that there were no pedestrians or cross-traffic present or approaching those intersections. Hence, there could be a reduced need for left-right head turns during brief stops at such intersections.

8. In any on-road driving study, university Internal Review Board protections ensure that drivers are protected from the conditions of a crash or near-crash, and generally require that drivers be advised that they are so protected. Hence, drivers participate with awareness that their safety is provided for at all times. Indeed, two experimenters were always the vehicle with the subject driver, and the front seat experimenter “notified the driver of any potential roadway hazards” [1, p. 22]. Because the driver knew they would be told of any potential hazards by the experimenter, then it is plausible that the driver would give a lower priority to head turns that any or all of the following tasks: a) following the lead vehicle, b) braking, c) steering, d) responding to the head-mounted red-green detection response task, e) performing the auditory-vocal tasks.

9. No analysis was made of actual glances, only head turns, making it misleading to label the head turn metric as “glances” as was done in the Utah study [1, Figure 26]. Some drivers may have chosen to make lateral glances with their eyes after a partial head turn, instead of making a full head turn. Such glances would not have been detected by the Utah methods, and so such trials would be incorrectly scored as a failure to turn the head.

10. It is not stated how many degrees of neck rotation were considered a head turn as opposed to no head turn. Because the video analysts who made judgments of the head turn amounts could see and hear what the task condition was, there is thus possible observer bias - e.g., the head turns of subjects performing tasks with assumed higher cognitive load could have been scored more strictly, resulting in fewer tabulated head turns.

1.3.5 Other Self-Regulatory Variables Not Mentioned in the Utah Study

A survey study [29] found that the majority of drivers “reported modifying their driving when engaged in other tasks. Most commonly, drivers reported reducing speed, pulling over to side of road, increasing following distance, and stopping their vehicle.” The survey also found that “In order to manage their exposure to risk, subjects reported that they will avoid engaging in distracting activities in a range of situations, including poor weather conditions, when driving at night, in heavy traffic, in unfamiliar areas, on winding or curved roads, and around schools.” Specific additional self-regulatory variables not reported in the Utah study, but were likely present there, are discussed next.
1.3.5.1 Speed

Speed data were not reported in the Utah study [1]. However, the mean following distance over a given 10-min task trial (Figure 3) varied in direct proportion to the Index (Figure 1), as shown in Section 1.3.2. It follows that the subjects must have reduced their speed relative to the pace car at one or more points in a given 10-min trial, or it would have been impossible for the subject vehicle to have increased its following distance relative to baseline as shown in Figure 3.

Previous studies performed by some of the same Utah investigators have shown that subjects do in fact reduce speed with auditory-vocal cognitive load [12, 13, 14, 15, 16, 30], a result which is consistent with virtually all other studies that used a speed metric [2, 24, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58]. Some may claim that reduced speeds associated with increased following distances cause an increase in crash risk. Although higher than average speeds are unquestionably linked with increased crashes, fatalities, and injury severity, the question of whether speeds lower than average increase crash risk has a controversial history [59, Chapter 8]. However, the most recent and carefully-controlled studies find no increased crash rate for vehicles traveling slower than surrounding traffic [60, 61, 62], after excluding crashes due to forced slowing down such as at intersections (for reviews see [59, Chapter 8; 63]).

1.3.5.2 Lane-Keeper

Lane-keeping metrics were not reported in the Utah study for its simulator experiment, even though these data are readily available in any simulator database. The attentional effects from cognitive load associated with auditory-vocal tasks give rise to improved lane keeping in virtually all driver behavior studies to date [45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63]. Improved lane-keeping during auditory-vocal tasks reduces crash risk, contradicting the Utah Index assumption of increased risk.

1.3.5.3 Shedding Tasks

Drivers are known to self-regulate by shedding other in-vehicle tasks while performing secondary tasks with attentional effects from cognitive load [27, 93, 94, 95]. However, drivers were expected to perform only the tasks assigned to them in the Utah study, so task shedding could not have been observed in the Utah study.

1.3.5.4 Gaze Concentration

The slight reduction in left-right head turns at an intersection during a task in the Utah study (Section 1.3.4.3) may be a result of “gaze concentration.” The formal definition of the gaze concentration metric is in terms of the percentage of glance time during a specified interval to a small specified area in the forward roadway [96, p. 143]. Hence, correct measurement of gaze concentration requires precise measurement of eye movements, which was not done in the Utah study. Increased gaze concentration during an auditory-vocal task is a well-established driver behavior [2, 27, 95, 97, 98, 99]. It can be conjectured that the decrease in left and right head turns in the Utah study may be indicative of gaze concentration, although this conjecture requires validation. The question of whether gaze concentration reduces or increases crash risk is also an active area of investigation. In general, gaze concentration means more “eyes on the road” time, which reduces crash risk, as per numerous naturalistic studies [e.g. 100].

Gaze concentration also reduces glances to the mirrors and to the speedometer, which usually tend to reduce crash risk [100]. Some may claim that gaze concentration thus increases crash risk. However, if drivers are self-regulating to minimize crash risk by staying in their lane [77], slowing down (Section 1.3.5.1), and choosing not to pass other vehicles during auditory-vocal tasks [2, 75], glances to mirrors or speedometer might not further reduce crash risk. Recarte and Nunes [97, p. 132] state, “Drivers can achieve an acceptable speed control with reduced speedometer inspection when they have no need to respect a particular speed limit restriction,” and “the relevance of mirrors inspection is strongly dependent of the traffic density and of drivers’ self-paced intentions to perform changing lane maneuvers.”

Interestingly, whenever a lead vehicle brake light event is detected during auditory-vocal tasks with associated gaze concentration, the gaze concentration ceases, and glance rates increase dramatically to areas other than the forward view [2, Figure 2-27]. Thus, auditory-vocal tasks (even in the presence of gaze concentration) put the driver in a better position to shift attention back to primary driving tasks than do visual-manual tasks, which exhibit no such increase in scanning behavior after hazard detection [2]. Thus, central vision stimulates a scanning of areas outside the vehicle (whether through mirrors or direct views) other than to the forward view, in order to better understand the trouble and/or how to avoid it.

Nevertheless, some still speculate that gaze concentration increases crash risk, despite improved eyes-on-road time. For example, one of the earlier Utah studies using similar simulator methods conjectured that gaze concentration during conversation tasks increases crash risk because of “inattention blindness” [12], a term often misused when applied to driving safety studies [101, 102]. Yet, auditory-vocal tasks, which increase gaze concentration, cause only slight increases in visual event RT vs. a “just driving” baseline [2, 103, 104]. The data shown in Section 1.3.6 below demonstrates that tasks with such slight increases in RT are not associated with increased crash risk.
1.3.5.5 Older Driver Self-Regulation

Only younger drivers were in the Utah study (18 to 30 years, with an average age of 22.2 years [1, p. 10]). It is likely that older drivers would score worse on many of the metrics used to create the Utah Index. However, studies [45] show that older drivers have lower crash risk than younger drivers. These two observations taken together contradict the Utah assumption that increases in its Index are associated with increased crash risk - the opposite is the case for older drivers.

For example, older drivers have about a 200 ms increase in mean RT to a red-green remote detection response task (DRT) during simulated driving with or without a cell phone conversation, which is longer than that for younger drivers [105]. There is an increased fatality risk for older drivers given that a crash occurs, but that is well-established to be due to increased physical fragility, rather than increased risk of crash involvement [106, p. 153]. Older drivers also have slightly more crashes per vehicle miles traveled than middle aged drivers, but that is due to low-mileage bias (older drivers drive fewer miles, but use more local roads where crash rates are higher). When the crash rate per distance traveled is estimated in age groups matched for yearly exposure, “there is no evidence for higher risk with increasing age” [107]. In short, older drivers in general have the lowest crash risk and further contradicting the Utah study assumption of “a monotonic relationship between cognitive distraction and crash risk” [1, p. 30].

1.3.6 Real-World Crash Risk and the Utah Index

The Utah assumption of a monotonic increase in crash risk for increasing cognitive load is contradicted by the relative risk (RR) estimates for cellular conversation (“Talk”) derived from NDS with objective measurements of calls, crashes, and driving times [4, 24, 108]. That real-world evidence shows that the RR of a safety-relevant event during Talk is no greater than baseline driving with no Talk, and may even decrease (a protective effect) [109]. Instead of citing these newer studies, the Utah report cites older case-crossover studies based on subjective crash and driving time estimates, which, due to parttime driving and call misclassification bias, substantially overestimated the Talk RR [109]. Figure 5 plots the Utah Index score for Talk tasks [1] on the horizontal axis, vs. 5 newer NDS Talk RR estimates [4, 24, 108] on the vertical axis.

The blue “No Talk” square in the upper left of Figure 5 shows that the Utah Index baseline value of 1 (“Single” in the Utah study [1]), has a corresponding RR of 1 (in the NDS studies [4, 24, 108]) by definition. As the Index increases from 1 to 2.25 for Hands-Free Talk, and 2.45 for Hand-Held Talk, the RR does not increase as predicted by the Utah monotonically increasing crash risk assumption. Instead, the RR estimates decrease to a range of 0.62 for OnStar Integrated HF Talk [4] to 0.79 for personal cell phone Hand-Held Talk [24]. Such decreased RRs directly contradict the Utah monotonically increasing crash risk assumption for their Index [1, p. 30], which, to be valid, requires all RR estimates in the study to be greater than or equal to the “No Talk” baseline value 1. In fact, the overall Talk RR estimate is 0.61 (95% CI 0.51 to 0.74), a protective effect, after adjusting the RR estimates from five major epidemiological studies for bias and pooling them [109].

The current study postulates that the reason for this protective effect is self-regulation by the driver. For example, the increased following distances found in the Utah data [1] compensate for the increases in brake RT during the different auditory/vocal tasks, as shown in Figure 4. Another plausible reason for the protective effect is an increase in driver alertness during Talk, which reduces drowsiness [110]. Using a baseline matched to crash/near-crash cases by driver demographics, time-of-day, and GPS location, the “moderately difficult” secondary task group (which includes cell phone conversation) reduced the drowsiness RR for crashes/near-crashes from 55 to 24 [110, Tables 5A, B].
1.4 Discussion

A re-analysis of the Utah study experimental data found evidence that for drivers in the Utah study, self-regulation from increased following distance provided an effective countermeasure against increased brake response times attributed to the attentional effects of cognitive load. Specifically, a re-analysis of the original Utah mean data found that, on average, drivers self-regulate their following distances to compensate for the slight increases in brake RTs while performing auditory-vocal secondary tasks, including interactions with machine voice systems as well as cell phone conversations.

The real-world crash data available for the hand-held and hands-free Talk tasks and the baseline single task indicate that the crash RR does not increase as the Utah Index score increases, but actually decreases (Figure 5).

1.4.1 Why Self-Regulation Is Not Commonly Observed in Experimental Studies

In experimental studies, subjects are required to perform the tasks under study in order to enable experimenters to measure the physical and/or mental demand of those tasks. As such, experimental studies cannot measure a subject's willingness to perform the task at any given time in real-world driving [45]. If the driver chooses not to perform the task, or slows down to increase following distance to the vehicle ahead while minimizing his or her eyes-off-road time, the experimenter would not be able to evaluate the potential attentional effects of the cognitive load presumably associated with the task under investigation. Such self-regulatory behavior is evident in NDS data (Section 1.3.4), and has been inferred from real-world crash data for cellular conversations [4].

However, experimental studies, if carefully designed, can measure the adjustments in driving behavior that subjects make when they engage in tasks such as cell phone conversation (although such studies cannot assess whether or when drivers would choose to perform such tasks while driving under naturalistic conditions). Such self-regulatory effects during task performance were evident in the current reexamination of the Utah study data, but were not noted in the Utah report [1].

The Utah report is based on experimental data (gathered in the laboratory, simulator, and controlled on-road experiments), and as such contains no naturalistic or real-world non-experimental driving. Therefore, there was no driver choice in the Utah experiments about the tasks they were asked to perform, or whether, when and how to perform them. The 32 subjects in the Utah simulator study were all university students, so they would be used to following instructions. In a typical experiment of this type, test subjects are required to perform tasks that can be made arbitrarily difficult in order to assess the possible attentional effects of “cognitive load.” Such experimental studies therefore do not assess how real drivers would behave when performing such tasks under real-world driving conditions. The requirement to complete the secondary task also applies an artificial time pressure to the subject, unlike self-regulation in real driving. Thus, experimental studies cannot account for effects of self-regulation by drivers, whereby drivers reduce their driving demand in order to perform the tasks and maintain their “virtual safety margin” [46]. The current paper shows that such self-regulatory behavior was evident in the Utah data during performance of auditory-vocal secondary tasks.

Therefore, the data in the Utah study fail to support any of its 3 main conclusions [1, p. 30]: (1) “there are significant impairments to driving that stem from the diversion of attention from the task of operating a motor vehicle” (2) “the impairments to driving are directly related to the cognitive workload of these in-vehicle activities” and (3) the “adoption of voice-based systems in the vehicle may have unintended consequences that adversely affect traffic safety.”

Specifically, it is shown that drivers in the Utah study increased their following distance in direct proportion to the cognitive index score of secondary auditory-vocal tasks. In fact, the Utah study subjects did so in such a way that their safety margins increased to compensate fully for the slight increases in brake response time to a lead vehicle applying its brakes during these controlled experiments.

1.4.2 Comparison with NDS

Experimental and naturalistic driving data may at times appear to give contradictory answers. Often this is because they are not measuring the same thing. It is therefore important for a researcher to choose the correct method for the research question at hand. An experimental study, such as that conducted by the Utah investigators, is arguably not the correct method to estimate crash risk. Its assumption that its results can be interpreted as predicting actual crash risk is speculative. In fact, the current study demonstrates that several of the variables in the Utah study Index are predicted to reduce crash risk rather than increase it. No purely experimental results can prove anything about crash risk, and in fact, equally valid arguments can be made that any given experimental result could either increase or decrease real-world crash risk. NDS results find that the amount of time the eyes are off the road during visual-manual tasks, both cumulatively [36] and via long single glances [111], are the chief safety concern, not cognitive distraction from auditory-vocal tasks wherein the eyes remain on the road.

NDS investigations to date have also undershot their potential for elucidating driver behavior, crash causation, and the role of driver self-regulation in mitigating crash risk. Looking at short video clips and brief epochs of vehicle performance data neglects the big picture of drivers’ tactical and strategic behavior, of which driver self-regulation is a key component. Examining a few seconds of video and vehicle data around the time of a crash or near-crash event cannot answer the question of what tactics and strategies were being used by drivers in a
given environmental and driver condition that got them involved in a safety-critical event in the first place. Moreover, just as important, what tactics and strategies were being used by drivers that kept them from being involved in crash and near-crash events? The fixation on crash and near-crash results when there are far larger amounts of non-conflict driving data that, if analyzed, could teach positive and protective driving behaviors (“what drivers do right”) that could “crowd out” behaviors that result in crash and near-crash conditions.

Similar questions arise regarding the analysis of the typical short duration (e.g., 6 s) samples of baseline driver and vehicle data in many naturalistic driving studies. True, having a proper baseline is necessary to understand driving safety. However, it is almost certainly a false assumption that a baseline video clip, however sampled, contains the same tactical and strategic behavior that was present around the time of a safety-critical event. A combination of driver trip analysis from key-on to keyoff [112] with naturalistic study methods for both safety-critical events and baseline driving, will be required to fully understand tactical and strategic driving behaviors in different environmental and road conditions, and their relationship to crash risk.

1.4.3 Driver Self-Regulation and Strategic Management of Multitasking
Driver self-regulation is part of the broader category of strategic management of task engagement during multitasking, which is controlled by the brain’s executive attention network [113]. Under normal driving conditions, the executive attention network exercises cognitive control of when a task is initiated (that is, under what conditions), and how much attention is allocated to the primary driving vs. secondary tasks at any given moment. For example, assume the driver decides to allocate cognitive resources to a secondary task. The data cited and analyzed in the current paper show that the typical driver might strategically reduce the load from the primary driving task by slowing speed slightly, lengthening headway slightly, and concentrating gaze on the forward view, which is the most important area for hazard detection and lane management. These added safety margins offset the reduced safety margin from the slightly increased RT in the dual-task situation. Auditory-vocal tasks typically offer more opportunity for strategic management than do visual-manual tasks, since eyes and hands need to be in continuous use for most of the time during visual-manual tasks. However, visual-manual tasks that are simple and well designed, with interruptible subtasks presented at a pace controlled by the driver, also offer opportunity for strategic management, more so than a text-heavy visual-manual task with an element of time pressure for completion.

1.4.4 Limitations
The re-analysis of the Utah data in the current study suggests that, when averaged over drivers, self-regulatory behavior compensates for distraction from auditory-vocal tasks for key driver performance variables in the Utah Index. However, some drivers may have poor self-regulation and hence are more crash prone. Thus, while the current study and others [8, 109] indicate that the attentional effects of cognitive load do not increase crash risk on average (likely due to appropriate self-regulation on the part of the typical driver), some individual drivers may fail to adequately regulate their behavior and thus increase their crash risk. Individual subject data, not just mean data, are needed to examine this question, and these were not published in the Utah report. Further studies are needed to explore individual differences in driver self-regulation and the strategic management of multitasking while driving.

1.4.5 Recommendations
Public and private funding sources for driving safety research spent on efforts intended to eliminate or ban the general use of voice interactions in vehicles will not reduce the number of crashes in the overall driver population. Further research funding aimed at reducing crash risk will be more productively spent by investigating methods to screen for drivers who do not self-regulate their driving appropriately, whether from drowsiness [110], alcohol, drug abuse, lack of driving experience, illness, or medical impairments. In short, it is recommended that efforts be made to improve self-regulation and strategic management of multitasking by drivers, rather than misguided efforts to curtail auditory-vocal tasks while driving, as implied by the Utah report [1].

1.5 Summary/Conclusion to Part 1
The assumption and implicit conclusion of the recent Utah experimental study that auditory-vocal tasks performed in the vehicle while driving increase real-world crash risk are not supported by its own or external data. Upon re-analysis, the Utah study’s data indicates that increased following distances offset the slight increases in brake response time during auditory-vocal tasks. In addition, the predictions of the Utah cognitive distraction index of elevated crash risk from cell phone conversations were not supported by evidence from naturalistic driving studies. In conclusion, the auditory-vocal tasks in the Utah study are not expected to cause impairments to operating an automobile when performed under real-world driving conditions.

PART 2. COMPARISON OF UTAH DATA TO CAMP-DWM DATA

2.1 Introduction
In Part 2, experimental data from the Utah study [1] are compared to the visual-manual and auditory-vocal tasks in the Crash Avoidance Metrics Partnership Driver Workload Metrics (CAMP-DWM) project [2]. The driver performance effect sizes arising from the auditory-vocal tasks in the Utah study [1] are compared to those of visual-manual tasks, using similar methods of testing. Specifically, Part 2 contrasts key variables and results between the Utah data [1] and those of the
CAMP-DWM project of Angell et al. (2006) [2]. These studies used comparable simulator and detection task methods, but the CAMP-DWM project encompassed a broader range of secondary tasks tested during driving, including visual-manual and auditory-vocal tasks in the same task set, tested using nearly identical methods. The comparison is useful, because it allows one to see how the Utah study results for auditory-vocal tasks compare with those for visual-manual tasks tested in a similar manner.

2.1.1 Objective

The objective of Part 2 is to compare some key results of the Utah study [1] for auditory-vocal tasks, to those of the CAMP-DWM study for auditory-vocal and visual-manual tasks [2].

2.1.2 Limited Tasks Lead to Narrow Viewpoint

While several forms of potential attentional effects from cognitive load were evaluated in the Utah report [1] (e.g., listening to the radio, talking on a cell phone, talking to a passenger, interacting with a speech-to-text system), the Utah investigators did not analyze a comprehensive set of common real-world tasks using their experimental protocol. Furthermore, they did not cite the experimental studies that have so analyzed a comprehensive set of common real-world tasks using a consistent experimental protocol. For example, Angell et al. [114] and Young and Angell [115] analyzed 78 real-world tasks in 5 vehicles using the PDT paradigm (now called Remote Detection Response Task (RDRT), the same as employed in the Utah simulator study) in both laboratory experiments and track studies. Angell et al. [2] analyzed 23 visual-manual and auditory-vocal tasks. The Utah study [1] did not cite these results, nor make any comparisons between its data and the data, results, and conclusions of these other studies - despite the fact that the Utah study [1] used similar experimental protocols to these earlier studies. Furthermore, the analyses of Young and Angell [115] and Young [111] have clearly established the sensitivities of the DRT methods to small gradations in the attentional effects of cognitive load due to visual-manual tasks. For these reasons, a comparison of the Utah results [1] with those from the CAMP-DWM study [2] with a broader range of tasks seems useful and warranted.

2.2 Methods

A detailed comparison of the methods employed by the two studies is required, because any differences in the results between the Utah and CAMP-DWM studies could be attributed to differences in the methods, rather than differences between auditory-vocal and visual-manual task properties.

2.2.1 Simulator

A fixed-based simulator was used in both studies. The STISIM simulator in the CAMP-DWM study [2] and the L3 Patrol-Sim simulator used in the Utah study [1] were different in some ways (e.g., single screen vs. multiple-screen; side-view mirrors in the Utah simulator but not in the STISIM; different simulated routes, traffic, and so forth). However, these are both simulator venues and little difference is expected in the RDRT response time variable of interest from differences in the particular type of simulator used. In both studies, subjects had to follow a lead vehicle in the simulator, and perform various secondary tasks. In the ISO DRT methodology [116], this simulator venue is known as the “dynamic” set-up [116].

2.2.2 Stimulus Protocols

The basic stimulus set-up in both studies was similar, now termed the Remote Detection Response Task (RDRT) [116]. There were some minor differences in the stimulus protocols.

2.2.2.1 CAMP-DWM Study DRT Stimulus Protocol

The DRT protocol in the CAMP-DWM study [2] used a computer-controlled, long-wavelength laser light that periodically flashed a spot on a screen in the left forward view of the subject. This stimulus set-up is traditionally known as the “PDT” for peripheral detection task, but is now termed the RDRT (Remote DRT) in the ISO DRT working draft [50]. Remote means the stimulus is displayed remotely from the body, whether in the peripheral or central field of view. The 1-sec duration light appeared red in color and came on randomly, repeated at between 2 to 10 seconds inter-stimulus intervals. A hit (i.e., correct response) was recorded if the response occurred between 200 and 2000 ms after stimulus onset (see Appendices to Angell et al. [117, p. D-14]), and a miss if not. The display screen also contained a driving scene projected from the simulator. The response time (RT) was averaged across the hit RTs during each task.

2.2.2.2 Utah Study DRT Stimulus Protocol

The Utah study’s version of the DRT task in the simulator was implemented by “mounting the red/green light on the vehicle dashboard directly in front of the participant” [1, p. 17]. It thus had the light mounted away from the body, and so falls into the general category of the RDRT, as was employed in the CAMP-DWM study. (Note that the Utah study did not use the head-mounted DRT version in its simulator study, where the stimulus is mounted on the head. The head-mounted DRT was used only in the Utah laboratory and road studies. The Tactile DRT [101, 116], which controls for the effect of eye movements on the DRT response times and miss rates, was not used in the Utah study.) The Utah study used the DRT working draft [116] timing of 3-5 sec inter-stimulus intervals with a 1-sec duration light stimulus. The Utah study did not use a single colored light as in the main body of the ISO DRT draft standard, but instead used a dual-color light, which could be either red or green. The subjects were told to respond to one of the hues (green) and not to respond to the other hue (red).

Because of the two hues, the Utah DRT stimulus is therefore not part of the main ISO DRT draft standard, but it is described in an informational Annex D in the current version of the ISO DRT draft [116], where it is termed the Dual Remote Visual DRT (DRV-DRT). The DRV-DRT was developed by Wayne
State University and has been used for some time in the driving safety literature (e.g., Young et al. [101, 118, 119, 120, 121], Hsieh et al. [52, 62, 63], Seaman et al. [122]), and was therefore agreed to be incorporated into Annex D in the draft ISO DRT standard [116]. (The current version of Annex D [116] also lists many references for this task in the cognitive neuroscience as well as psychology literature, where it is commonly known as the "Go/No-Go Task.") The Utah DRT stimulus differs in several respects from the ISO informational annex D method. First, it requires subjects to respond to the green light, and not respond to the red light, the opposite of the method in Annex D. The response to the red light in the Wayne State set-up was intended to mimic a braking response to a brake light illumination on a forward vehicle, or the response to a traffic control red light. Previous work has also shown that the response times to a red light in the red/green DRT method in Annex D are equivalent to response times to a single red light (without a green "no-response" light) (Young et al. [101]). In other words, the red-green method in Annex D gives rise to the same RT and miss rate result for the red light as the single red light RDRT implementation in the CAMP-DWM study [2], or the ISO DRT draft standard main body RDRT. The green light is intended in the Wayne State studies to measure the ability of subjects to inhibit their responses to non-target stimuli (that is, to not be distracted by irrelevant stimuli), a quality equally as important for reducing driver distraction as paying attention to relevant stimuli. The Utah study did not use their two-color stimulus to investigate behavioral inhibition.

Another major difference of the Utah methods from the ISO DRT Annex D methods is that the green target light in the Utah study (to which the subjects were to respond with a finger press of a button) was present in only 20% of the events (the other 80% of the events were the red lights, to which the subjects were told not to respond). Thus, the effective inter-stimulus interval in the Utah study was 15 to 25 sec, compared to 3-5 sec in the ISO DRT draft standard [116], and the 2-10 sec used in CAMP-DWM study [2]. The Utah stimulus to which subjects should respond would therefore be less predictable (because of the greater time uncertainty), which would lead to unknown differences in the size of the attentional effects of cognitive load, compared to the effect sizes using the ISO DRT draft standard methods.

In summary, the results from the Utah simulator study [1] can be compared to those from the CAMP-DWM simulator study [2], because the simulator scenario and the DRT methodology that were used, despite some differences, should give similar results at least for the simple red light RDRT method used to measure the attentional effects of secondary tasks.

### 2.2.3 Secondary Tasks

The major difference between the studies is that the CAMP-DWM study [2] used a wider variety of task types, because it included both auditory-vocal and visual-manual tasks. The Utah study [1] focused on just auditory-vocal tasks. Although the Utah approach has surface validity, it suffers from loss of context. It is preferable to use a wide range of different tasks and task types, with varying types and degrees of attentional load. This greater range places task effects in a larger context that allows for more meaningful comparisons among tasks and scenarios. For example, using only a small set of tasks or only tasks of one type might give the false impression that there are major differences between the tasks, when the results for the limited set of tasks are examined under a “microscope.” A wider view with a greater variety of tasks might show that the apparently large differences are actually quite small. While a wider view provides more context, it requires collecting data on a wider set of tasks of varying types, and is thus more resource-intensive than collecting data on just a few tasks, or just auditory-vocal tasks, as was done in the Utah study [1].

The “Just Drive,” “Book on Tape Listen,” and “Biographical Q&A” tasks (which are representative of a hands-free phone conversation) were essentially the same in the two studies, but were administered in different ways. For example, “Just Drive” in the CAMP-DWM testing was 2 min vs. 10 min in the Utah study. Each trial of the CAMP-DWM auditory-vocal tasks were designed to last roughly 2 min, while each trial in the Utah study lasted 10 min (tasks were repeated to fill out the 10 min where necessary). Repeating the same task over and over again in rapid succession in the same trial is obviously not natural task behavior, whether driving or not, and has been recommended against [123, 124], although see [125]. (This repetition of the same task was specifically to be avoided in the Alliance guidelines [118] testing protocols, for example.) Whether it produces similar driver performance behavior to natural driving is an area for investigation. Repeatedly doing a task would be expected to lead to task performance improvements (and therefore less load) from more practice, but could also lead to task and driving performance decrements due to fatigue, leaving the result uncertain, compared to use of a more naturalistic single-task test procedure in a given trial. In addition, test subjects probably perform tasks differently if only expected to do it once, vs. when they know they will have to start the same task all over again as soon as they finish. For example, their motivation to be fast and efficient may decrease in the latter case.

However, a control for these protocol differences is already built into the comparison of the results across the two studies [1, 2]. Namely, if the auditory-vocal tasks in the two studies give rise to similar results, then there is some evidence that the protocol differences are small enough that they do not produce large differences in the results for auditory-vocal tasks.

Therefore, any differences between the Utah auditory-vocal tasks and the CAMP-DWM visual-manual tasks would be expected to arise from differences in auditory-vocal vs. visual-manual task types, and not from protocol differences in the test methods in the two studies.
In sum, to address the narrow task range in the Utah study [1], Part 2 has placed the auditory-vocal task data from that study into a wider context by looking more broadly at the larger set of tasks in the CAMP-DWM study that includes both auditory-vocal and visual-manual tasks [2]. The Utah study simulator RTs were digitized from [1, Figure 4, p. 36].

2.3 Results

Figure 6 compares the DRT RTs for the auditory-vocal tasks in the Utah simulator study [1] (left grey bars), vs. the auditory-vocal tasks in the CAMP-DWM simulator study [2] (center green bars), vs. the visual-manual tasks in the CAMP-DWM simulator study [2] (right blue bars).

Figure 6 shows that, in general, the 500-700 ms range of DRT RTs for the CAMP-DWM auditory-vocal tasks (center green bars) was well replicated by the Utah auditory-vocal tasks (left grey bars). This result shows the robustness of the ISO DRT method [116] to measuring the attentional effects of cognitive load for auditory-vocal tasks, despite the differences in methods, subjects, simulators, etc. between the two studies described in Section 2.2 (e.g., red/green lights vs. single red light, different inter-stimulus interval ranges, different stimulus location, participants, inter-stimulus times, repetition of tasks, etc.). For the three specific auditory-vocal tasks that were essentially the same between the Utah and CAMP-DWM studies (Single* vs. Just Drive*, Book-on-Tape Listen†, and Hands-Free‡ vs. Bio Q&A‡), the reported RTs are in the same range as is apparent from inspection of the matched tasks in the grey and green bars in Figure 6.

Further evidence of similarity for the auditory-vocal tasks in the two studies [1, 2] is the finding that all of the RTs increase for the triple-task condition (i.e., a secondary task plus the driving and DRT) compared to the dual-task condition (i.e., “Just Drive” plus the DRT). This similarity is expected because of the additional attentional load required for switching between three tasks instead of two, and not necessarily to the amount of cognitive load arising from the secondary task itself. An added attentional effect would happen with any third task, regardless of its associated cognitive load.

The main objective of Part 2 is to compare the RTs for the auditory-vocal tasks in the two studies, to the RTs for the visual-manual tasks (right blue bars) in the CAMP-DWM study. These represent the DRT RTs for visual-manual tasks using the identical procedures used for auditory-vocal tasks in the CAMP-DWM study. Without exception, every visual-manual task in the CAMP-DWM study (right blue bars) had longer DRT RTs than all of the auditory-vocal tasks in both the CAMP-
DWM (center green bars) and Utah studies (left gray bars), as is obvious by visual inspection of Figure 6. Simply stated, this means that every visual-manual task tested in the CAMP-DWM study [2] (including some common tasks considered socially acceptable while driving, such as searching for coins in a cup holder to pay a toll, or manually tuning a traditional radio with knobs, etc.), has more attentional effects than any auditory-vocal task tested in either study.

Visual-manual tasks have more eyes-off-road time as shown in the CAMP-DWM study [2] glance data for those same tasks when performed on the road or track (the glance data for the simulator was not tabulated in the CAMP-DWM study). There is thus an additional explanation for the longer RTs for the visual-manual tasks, because they would include the attentional effects arising from visual load as well as cognitive load [47]. However, the point of the comparison in Figure 6 is to show that the RT effects (as measured by the DRT) from auditory-vocal tasks that are deemed to be “impairing” the driver in the Utah report [1], are far smaller than the RT effects from visual-manual tasks that are commonly accepted in the vehicle today, regardless of the origin of the effects. A similar observation was made about auditory-vocal being safer than visual-manual tasks in the CAMP-DWM report [2].

There are currently no criteria for DRT response times in the final NHTSA visual-manual driver distraction guidelines [126]. However, there was a 1-s DRT RT criterion set in the original proposed NHTSA guidelines [127] (which was left out of the final NHTSA guidelines [126]). It is noteworthy that all tasks in Figure 6 meet the 1-s proposed NHTSA DRT criterion [127]. NHTSA [128] is on public record as stating it intends to come out with an extended set of Guidelines in 2014 that will include auditory-vocal tasks within its scope. It is quite possible that these extended Guidelines will include one or more of the ISO DRT methods, as these methods are now deemed robust enough to be codified as a draft ISO standard (ISO DIS 17488) [116]. This draft ISO standard [116] has no acceptance criterion for response time, as it is intended as a procedural standard and not a criterion standard. However, if NHTSA adopts the DRT in its auditory-vocal guidelines scheduled to be issued in 2014, it will likely set an acceptance criterion for DRT RT, as it did for visual-manual tasks using other metrics. The analysis in Part 2 shows that if the eventual criterion were set at the level of the Utah study's Category 5 OSPAN task (the highest, most cognitively-loading level on their hurricane-like scale), and applied to visual-manual tasks as well as auditory-vocal tasks, then all visual-manual tasks in the vehicle - including those that are widely practiced in everyday natural driving and are deemed to be socially acceptable, as well as driving-related tasks such as mirror checks and reading exterior trip-relevant signage - would not meet that criterion. Setting the criterion at “lower” levels of the Utah scale would only make the criterion even more stringent, and thus even less appropriate.

Note that the proposed NHTSA guidelines [127] (but not the final version [126]) allowed a comparison against the radio tuning task as a criterion, which is the same “RadioHard” tuning task used in the CAMP-DWM study (and the same radio tuning task that was used in the Alliance guidelines [118] as a reference task). Note that all auditory-vocal tasks in both the CAMP-DWM [2] and Utah [1] studies have a lower DRT RT than the RadioHard tuning task.

The current analysis suggests that it would make more scientific sense to set a general RT criterion for auditory-vocal (and visual-manual) tasks based on either the 1-s proposed NHTSA [127] DRT RT criterion, or visual-manual radio tuning, rather than anything connected with the Utah Index.

### 2.4 Discussion

The CAMP-DWM study [2] used a simulator, with a remote detection response task (remote DRT), as did the Utah simulator study. The CAMP-DWM study investigated some of the same auditory-vocal tasks as the Utah study (for example, “Radio listen,” and a “Bio Q&A” task as a surrogate for cell phone conversation). It also included visual-manual tasks prevalent in everyday driving, such as conventional radio tuning or searching for coins to pay a toll, which have a crash risk that is considered acceptable to the general driving public. All the Utah and CAMP-DWM auditory-vocal tasks scored better than these visual-manual tasks on the RT metric. That is, auditory-vocal tasks have a lesser effect on the orienting attention network that handles responses to visual events [129] than do visual-manual tasks.

The fact that visual-manual tasks are more demanding than auditory-vocal tasks [2] does not mean that all visual-manual tasks are “unsafe,” because “safe” does not mean zero task load. All tasks performed in a vehicle, even tasks related to primary driving such as checking mirrors, have task load. The claim in the Utah study [1] that the attentional effects of cognitive load from auditory-vocal tasks cause significant “impairments” to operating a motor vehicle for most or all drivers is not supported by comparison to visual-manual task loads.

### 2.5 Summary/Conclusion to Part 2

Part 2 directly compared the response times of the tasks in the Utah report [1] to the response times in the CAMP-DWM study [2]. It found that the auditory-vocal tasks about which the Utah report raises major concerns have response times that are below the response times of visual-manual tasks that are well within the boundaries of tasks considered acceptable in normal driving by the general public. They are also well below the 1-s response time criterion in the initial proposed NHTSA visual-manual guidelines [127].
In summary, the Utah study’s [1] claim that auditory-vocal tasks performed in the vehicle while driving cause serious impairments to driving performance is not consistent with the more comprehensive results from the CAMP-DWM study [2]. The latter study found greater RTs for visual-manual tasks, including several that are considered acceptable by the general driving public, than the auditory-vocal tasks impugned in the Utah report. As shown in Figure 5 in the current paper, the Utah Index [1] is associated with reduced real-world crash risk for Talk tasks, and therefore Talk tasks do not represent “impaired” driving as claimed by the Utah study [1]. Part 2 shows, consistent with that observation, that a key variable in the Utah Index (driver response time during a DRT), shows smaller attentional effects from auditory-vocal tasks than visual-manual tasks when performed in an automobile while driving, consistent with prior results [2].

In conclusion, auditory-vocal tasks such as those in the Utah study are not expected to increase crash risk or impair real-world operation of an automobile, compared to normal baseline driving without performing such tasks.

CONTACT INFORMATION
Richard A. Young
ryoun@med.wayne.edu

ACKNOWLEDGMENTS
I thank Bruce Papazian, Katja Kircher, and several anonymous reviewers for helpful comments, Amanda Zeidan for research assistance, and Sean Seaman for editing assistance.

DEFINITIONS/ABBREVIATIONS
CAMP-DVM - Crash Avoidance Metrics Partnership Driver Workload Metrics
D - Distance behind pace vehicle in simulator
DRT - Detection Response Task
DRV-DRT - Dual Remote Visual DRT
OSPAR - Complex series of math and verbal problems
PDT - Peripheral Detection Task
RDRT - Remote Detection Response Task
RT - Response Time
S - Speed of subject vehicle in simulator
T - Time behind pace vehicle in simulator

REFERENCES


