Voice Interaction Design

The Journal of AVIxD Association for Voice Interaction Design

Vol. 2, Issue 1, August 2017 pp. 1-34

Driven to Distraction? A Review of Speech Technologies in the Automobile

Richard A. Young
Founder and President
Driving Safety Consulting
5086 Dayton Dr.
Troy, MI 48085-4026
USA
richardyoung9@gmail.com

Jing Zhang
Founder
AutoSimpler
1366 Kensington Ave.
Grosse Pointe Park, MI
48320
USA
jz@autosimpler.com

Abstract

Speech technologies hold the promise of improving driver performance for many visual-manual secondary tasks, by enabling eyes-free and hands-free interactions. Unfortunately, speech interfaces have enjoyed only incremental growth since the early 2000s in the automotive industry. Instead, *mixed-mode* interfaces (speech combined with visual) have become increasingly common, and visual-manual interfaces are still dominant. This paper provides a historical overview of speech driver interface studies, including formal testing on a 2014 Toyota Corolla production vehicle, and a new analytical evaluation of the Apple CarPlay interface in the 2016 Cadillac ATS. Results indicate that eyes-free and hands-free speech (i.e., "pure" speech) improves driver performance vs. mixed-mode interfaces. Also, mixed-mode improves driver performance vs. "pure" visual-manual, for the tasks tested. The visual component of the mixed-mode and visual-manual interfaces increases off-road glances, a safety decrement. We recommend that future in-vehicle speech interface products should sensibly limit visual displays from showing information redundant to that provided by the speech interface. In general, we recommend pure speech driver-vehicle interfaces for secondary tasks wherever possible.

Keywords

Speech interfaces, driver performance, visual-manual, auditory-vocal, mixed-mode

1 Introduction

Conversational speech-only user interfaces have been of interest for decades. In science fiction, two of the first examples may have been the simulated natural language computer voice from the original Star Trek TV series from 1966 to 1969[1], and the “HAL 9000” computer natural language voice in the epic 1968 movie “2001”.[2] The first fictional voice system in a car may have been the KITT (Knight Industries Two Thousand) natural language voice system, the “Knight Rider” television series broadcast from 1982 to 1986,[3] where an actor covertly carried on the conversation for the car using a script.

However, what was science fiction yesterday sometimes becomes reality today, at least in prototype. One concept car that could control vehicle functions with speech was called the Quiet Servant, to be discussed in a later section. Volvo and Microsoft collaborated on a prototype system which recognized voice commands from a watch (such as “start engine”) and transmitted them to a car.[4] A prototype of a speech interface system that allows drivers to consult manuals while driving was also developed.[5]

The technology with which to implement an actual conversational speech-only user interface for automobiles has actually been available for almost two decades. However, going from a concept car to production has been resisted at executive levels, setting back speech interfaces in production vehicles, to be discussed in a later section. This review will summarize evidence that automotive speech user interfaces have substantial, measurable driver performance advantages over the visual-manual alternatives, as well as excellent customer acceptance when well-implemented. The Society for Automotive Engineers (SAE) has recognized the potential benefits and importance of speech technology in vehicles, and has accordingly developed guidelines for speech interfaces in a driver-vehicle interface (SAE J2988_201506, 2015).

Following Young (2014b), Young et al. (2017b), and the common use of these terms in the automotive industry, we define three “modes” by which tasks can be performed via an automotive driver-vehicle interface:

1. *Visual-manual mode* – A task is accomplished via screen displays and button presses, with no voice output from the system or vocal commands by the driver. Visual-manual tasks may still include tactile or auditory feedback (such as a “click” sound) from presses of a physical switch “hard” button, presses of a touch screen “soft” button, or gesture recognition. Visual-manual tasks are neither eyes-free nor hands-free. If the visual-manual mode has no other modalities it is here termed a *pure* visual-manual interface mode.
2. *Auditory-vocal mode* – A task is accomplished via speech input and audio (speech and otherwise) output, without visual information presented. To meet the SAE J2972_201403 (2014) definition of hands-free and eyes-free, an auditory-vocal interface cannot have more than one button press; e.g., a “voice button” press to activate the speech recognition system. If the auditory-vocal mode has no other modalities it is here termed a *pure* auditory-vocal interface mode.
3. *Mixed-mode* – A task is accomplished via auditory-vocal and some combination of visual and manual modes in the same task. The task is generally accomplished primarily by driver speech commands, but visual information is also presented, so the task is not “eyes-free.” If more than one button press is required to complete a mixed-mode task, the task is not hands-free according to the automotive definition (SAE J2972_201403, 2014). Mixed-mode interfaces are also sometimes referred to in the driver-vehicle interface literature as “multi-modal” interfaces (e.g., Mehler et al., 2015a).

[1] Retrieved from https://youtu.be/NM4yEOdIHnc
[2] Retrieved from https://youtu.be/HwBmPiOmEGQ
[3] Retrieved from https://www.youtube.com/watch?v=Boqh3mEawyU
[4] Retrieved from http://video.dailymail.co.uk/video/1418450360/2014/12/1418450360_3941102969001_blue-eye.mp4
[5] Retrieved from https://youtu.be/IwVN1deb9jM?list=FLVF4tIopoeOR8__U93Q_JRKw

Is driver performance better with a pure auditory-vocal interface than with a pure visual-manual interface for many secondary tasks? For example, do drivers stay in their lanes better and make fewer driving errors? If so, does a pure auditory-vocal interface have good customer acceptance? Both driver performance improvements and customer acceptance are necessary for success, because if an auditory-vocal interface worsens driving performance, then even if it had excellent customer acceptance, automakers might be reluctant to implement it. Conversely, if a pure auditory-vocal interface improves driver performance, then it would benefit society only if customers accepted it and purchased vehicles with such an interface.

There are as yet no naturalistic or real-world driving studies with crash data that allow the safety question of a generic voice interface to be directly answered. Therefore, the safety question is not the main focus of this review, which is limited to driver performance. In general, however, improved driver performance would be expected to give rise to improved safety. However, reduced driver performance in an experimental test does not necessarily translate to reduced safety, because of driver self-regulation (Young, 2014a,b; 2015a). Hence, to answer the safety question definitively requires analyses of crash data that remove as many biases as possible (e.g. Young and Schreiner, 2009; Young, 2015b,c; 2017a,b, in press). The safety question is briefly addressed in Section 6.2 of this paper under future research.

The main question addressed in the current review is therefore, "Are voice commands better for driver performance than button presses?" Specifically, we review driver performance results in several experiments that have directly compared pure visual-manual vs. pure auditory-vocal interfaces, and sometimes mixed-mode interfaces as well. In some of these studies, drivers performed the same task in the same vehicle, but with the different interface modes.

Note that acronyms have been minimized throughout this review, and the few that are used are defined in an abbreviations/definitions section at the end of this review. The paper is written based on the direct experience of the authors in the automotive industry, so a number of terms are used that may be familiar to those in the automotive industry but may not be familiar to those in the speech field. We have therefore tried to define these terms wherever possible for a more general audience, and give a list of acronyms and definitions near the end of this review. Please note also that space did not permit the providing of the full details of every study reviewed here, but citations are provided to all the studies, and the interested reader can refer to the cited references for further detail. Please note that this review is also not meant to be exhaustive of all the dozens of studies that have been conducted on automotive speech, but instead reflects studies that the authors have direct experience with, or otherwise believe are particularly important for speech interface development for automobiles, from their viewpoint.

Note that the overall content of this review of automotive speech interfaces is organized according to a historical timeline.

- Section 2: Early Research: 1999-2006
- Section 3: Recent Research: 2007 to 2016
- Section 4: New Research
- Section 5: Review Limitations
- Section 6: Future Research Recommendations
- Section 7: Summary and Recommendations
- Section 8: Conclusions

2 Early Research on Speech Interfaces: 1999-2006

2.1 Comparison of Pure Auditory-Vocal vs. Pure Visual-Manual Interface Modes

Young (1999, 2001a) measured and compared driver performance during secondary tasks using pure visual-manual vs. pure auditory-vocal interface modes in four devices in three vehicles. Participants were randomly selected from mid-size vehicle owners in the Detroit area except for age and sex balancing. The final 94 participants were: 50% under 50 years, and 50% over 50 years old, with no one over 65; 45% male participants and 55% female participants. The selections were random for technical literacy.

2.1.1 Test Set-up

The vehicles were electrically powered but stationary during the test. Secondary tasks were performed under "surrogate" driving conditions intended to mimic driving (Figure 1).

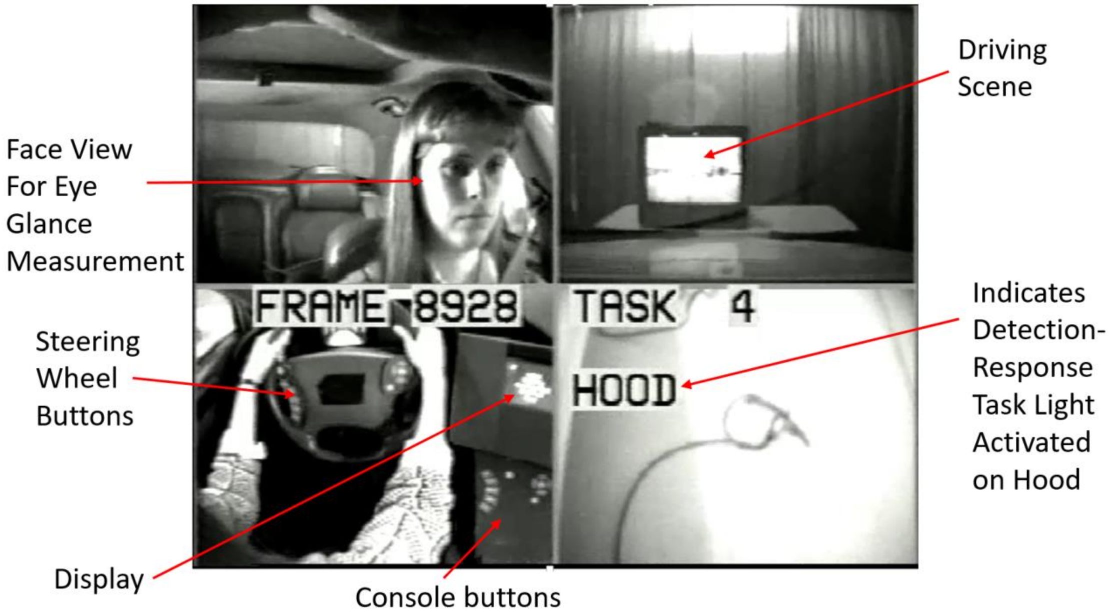

Driving Scene
Face View For Eye Glance Measurement
FRAME 8928
TASK 4
Indicates Detection-Response Task Light Activated on Hood
HOOD
Steering Wheel Buttons
Display
Console buttons

Figure 1. Driver and prototype vehicle set-up for testing pure auditory-vocal vs. pure visual-manual driver-vehicle interface modes

A video monitor in front of vehicle (lower left of Figure 1) played a continuous loop of a videotaped driving scene. A small red L.E.D. light was in the forward field of view of the driver, at the bottom of the monitor. The light was turned on at random times to mimic an activation of brake lights in front of the driver. The visual angle of the L.E.D. was not deliberately selected to be equal to the visual angle of the brake lamps in vehicles in the video, but coincidentally it was similar. The size of the light was identical for all task conditions so it could not affect the relative detection rate and response time for the different tasks. The brake pedal of the test vehicle was wired such that when the driver pressed it, it turned off the L.E.D. light on the forward scene, and recorded the brake response time to a data file.

This test method was known at the time as the "peripheral detection task" or what is now called the remote detection-response task (RDRT) in the International Standards Organization (ISO) standard 17488:2016 (2016), although the ISO standard uses a finger button press instead of a brake pedal push. Although the absolute response times are slightly longer for foot vs. finger (because of the longer length of the axons to the foot than to the hand, and the greater weight of the foot relative to the hand), the relative response times for foot and hand are the same across different task conditions.

Three tiny video cameras were mounted inside the vehicle and their video outputs were time-multiplexed and time-synchronized into a single image in one video recording (Figure 1). Camera 1 focused on the participant's face (upper left of Figure 1), to record glances. Camera 2 focused on a participant's hands (lower left of Figure 1), to record hands-off-wheel and hands-on-device behavior. Camera 3 focused on the forward view (upper right of Figure 1) to record the driving scene on the video monitor, and the illumination of the (simulated) brake light. The words 'Hood' and 'Brake,' to indicate the simulated forward brake light and the braking responses by the participant (lower right of Figure 1), were recorded on the same video, for later analysis if needed. A microphone was mounted on the vehicle interior to pick up vocalizations of participants and the voice output of the system under test (which were recorded on the same videotape as the camera images).

2.1.2 Tasks

The test participants performed two in-vehicle communication and navigation tasks ("navigate to school" and "call your daughter"). Conventional driving tasks were used as a baseline control: adjusting the left rear-view mirror, adjusting vents, and setting the cruise control. A given communication or navigation task was performed with the same start and end points whether performed with pure visual-manual, mixed-mode, or pure auditory-vocal interfaces. All tasks were performed by all participants for a given system. Of the four systems, three were commercially available: AutoPC (Microsoft, 1998); a 1999 Honda Acura navigation system; the Portico voice communication system.[6] The fourth was a prototype system specifically designed and developed for this test. The prototype system had a pure auditory-vocal mode, and two pure visual-manual mode, one with steering wheel buttons and one with console buttons (see Figure 1, lower left quadrant).

2.1.3 Training Procedure

Each participant was given individual prior instruction in the secondary tasks with storyboards. They then sat in a vehicle, and adjusted seating position. They were told to watch and steer to the roadway scene on the video monitor in front of their car. They were also instructed to watch for the small red light, and then make a quick tap on the brake when they detected it. Participants were instructed that their primary task was watching the driving video while keeping their hands on the wheel and eyes on the road as much as possible, along with responding with a pedal press to the small red light below the forward driving scene. An in-vehicle demo of the task by the experimenter was followed by in-vehicle practice until the participant could perform the task once without error and felt comfortable doing so. The tested systems included error recovery paths within each system. Participants were not specifically trained on these error recovery paths, but they learned about them during their regular practice trials. The trial was counted as successful if the participant reached the end goal, even with an error recovery if necessary.

2.1.4 Test Procedure

All experimental conditions were tested in a within-subjects experimental design. All devices, interfaces, and tasks were tested by all participants, with a randomization of the order in which each participant was tested. Participants performed the secondary tasks in all four systems, while also detecting the light and tapping the brake pedal to respond.

2.1.5 Surrogate Driving and Task Performance Metrics

The metrics for the secondary task itself were task completion time, the number of button presses to complete the task, and a count of the number of voice commands the participant required to do the auditory-vocal mode of a task. The percentage of "successful" trials, in which the given task was successfully accomplished by achieving the goal, and the percentage of "unsuccessful" trials, in which the driver failed to complete the task, were also tabulated. Two key driver performance metrics were response time and miss rate to the simulated brake light, which are a surrogate for crash avoidance. Other driver performance metrics were the glance measures of total eyes off road time, number of glances, and mean single glance duration.

One subjective metric was "workload rating" averaged across three scales; e.g., "How much mental and perceptual activity was required?" (1 = Easy, 100 = Hard). Another subjective metric was rating of "situation awareness," "How aware were you of surrounding traffic when you were performing the secondary task" (1 = Low, 100 = High).

2.1.6 Results

The results showed that pure auditory-vocal tasks, whether the trial was successful or unsuccessful, had better driver performance than pure visual-manual tasks in terms of reduced eyes-off-road time and quicker driver responses to visual events. The auditory-vocal mode took longer to complete than the visual-manual mode for the same task (i.e., a task with the same

[6] Retrieved from http://www.liquisearch.com/general_magic/later_developments

start and end points), but the longer task time was of no detriment to the other driver performance metrics examined here. That finding supports the decision of the SAE to limit the scope of its task time metric commonly referred to as the "15 second rule" (see SAE J2365_201607, 2016) to visual-manual modes only. This best practice specifically excludes auditory-vocal modes from its 15-s task time criterion.

2.1.7 Predictability of On-Road Test Results

Later research shows that the "surrogate" driving test method pioneered in this 1999 experiment (Young, 1999, 2001a) has high predictive ability (about 0.9 correlation) with driver performance metrics such as response time and miss rate to detection response task events in closed-road driving tests (Angell et al., 2002; Young and Angell, 2003; Young et al., 2005). In addition, this simple low-cost surrogate driver performance test using a driving video is also predictive of the response time and miss rate results from simulator and closed-road driving assessment of secondary tasks (ISO 17488:2016, 2016, Annex E).

2.2 Quiet Servant Concept Vehicle

As a consequence of the study in Section 2.1 and others showing the benefits of speech interfaces for reducing driver distraction, the vehicle program team for a 2003 production vehicle developed a "mule" vehicle using an auditory-vocal interface. (A "mule" vehicle is an existing production vehicle that has been converted to a pre-production concept vehicle.) This mule vehicle was designed, built, and tested from the years 2000-2002. It was a novel driver interface design, in that it had a clean dashboard with no displays and controls, apart from a high-quality analog clock in the center (Figure 2).

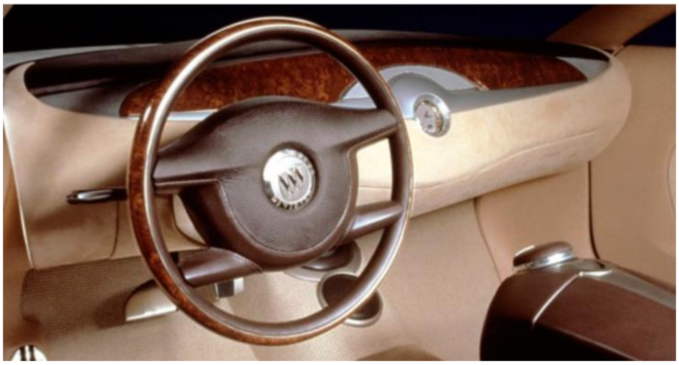


Figure 2. Quiet Servant interior. Photo from public sources[7]

The driver interface was mainly voice-based for all secondary tasks (e.g., radio tuning, climate adjustment), and many primary driving tasks (e.g., adjust mirrors), except for acceleration, braking, steering.

The instrument displays were replaced by a color head-up display (HUD) that itself was controlled by voice commands. A HUD allows drivers to keep their eyes on the road while viewing. The HUD provided vehicle information on the windshield at a virtual distance of 6 feet (i.e., the image was seen by the driver as floating near the front of the car). The image was transparent so it did not block the view of the roadway.

The early prototype depicted in Figure 2 had a "jog shuttle" (described by Buick as being a "mushroom-shaped combination mouse and joy stick"), located at the front of the center armrest. This control was replaced in later versions of the prototype vehicle with manual steering wheel controls (not shown in Figure 2). These controls were redundant to the voice controls (that is, drivers did not have to use them unless they wanted to). They allowed drivers to keep their hands on the wheel while performing manual operations if they wished to. The

[7] Retrieved from http://autosofinterest.com/2012/08/18/2000-buick-lacrosse-concept/2/

redundant steering wheel controls would also have allowed persons with hearing and/or speech impediments to use the interface while keeping their hands on the wheel.

The prototype was dubbed "Quiet Servant" (even though the vehicle spoke) because the driver performance tests found that the interface design minimized driver distraction to an extent that had not been previously seen in an automotive driver-vehicle interface. Indeed, in driver performance assessments in a static set-up in a garage, and in dynamic testing while driving on a closed road, the Quiet Servant had the best driver performance (i.e., the least amount of eyes-off-road time and the fastest responses to external visual events) of any automotive or manufacturer portable device product supplier in the lead author's experience. The Quiet Servant concept vehicle, both in bench tests, stationary test sites, and actual driving tests, also received excellent customer acceptance in numerous market research clinics (personal observation).

Unfortunately, Quiet Servant met with senior executive leadership resistance due to the radical design of the cockpit. A key senior executive sat in the vehicle for the first time and asked in apparent shock, "Where are the buttons and gauges?" The executive cancelled the program the next morning based on this personal opinion. This executive has stated in public speeches and in print (Lutz, 2011) that cancelling the Quiet Servant program was a "highlight of his career." In our opinion, this "management by subjective opinion," rather than by the evidence and data from driver performance tests and market research clinics, almost certainly retarded the development of speech interfaces for in-vehicle tasks at this particular major automotive company by many years. When the re-designed vehicle was finally released in 2005 after a two-year production delay because of the cancellation of the prototype, it had a conventional instrument cluster and center stack (Figure 3), conforming to this executive's traditional design direction.

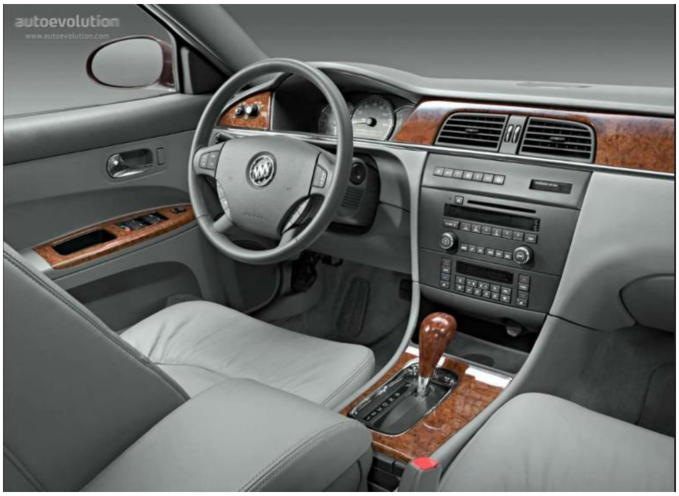


Figure 3. 2005 production vehicle interior with traditional instrument panel and conventional visual-manual display and control interface

2.3 CAMP-DWM Study

The "Crash Avoidance Metrics Partnership Driver Workload Metrics" (CAMP-DWM) program was a large multi-year driver performance study of secondary tasks. It was co-sponsored by NHTSA and Ford, GM, Nissan, and Toyota. It compared the driver performance of 23 auditory-vocal and visual-manual tasks in simulator, closed-track, and open road tests, and published its final results in 2006 (Angell et al., 2006a,b). This study found that on most driver performance measures evaluated, drivers performed better with auditory-vocal than visual-manual interfaces.

The exception was task time, where the selected auditory-vocal tasks typically took longer than the selected visual-manual tasks. However, task time is not relevant voice tasks, because there is no decline in other measures of driver performance for long voice tasks, unlike for visual-manual tasks (Schreiner et al., 2004a,b; Schreiner, 2006). That is why voice tasks are excluded from the SAE measurement methods for task time (SAE J2365_201607, 2016). See Appendix Figure A1 for an example for DRT response time, a measure of the attentional effects of cognitive demand (ISO 17488-2016).

See Figure A1 in Appendix A, which indicates that the attentional effects of cognitive demand arising from visual-manual tasks are greater than those arising from auditory-vocal tasks, contrary to common assumption.

3 Recent Research on Speech Interfaces: 2007-2016

This section reviews more recent research that confirms or rejects the conclusions in Section 2; namely, that pure speech interfaces give rise to better driver performance than pure visual-manual interfaces for secondary tasks (at least for those tasks so far tested). Several recent studies on mixed-mode interfaces are also reviewed.

3.1 Strayer et al. Studies of Speech Interfaces

A series of studies by Strayer and colleagues reached the opposite conclusion of the studies in Section 2. Strayer et al. (2013, 2014, 2015a) claimed that what they call "cognitive distraction" from speech tasks in vehicles causes substantial safety risks. They created what they called a "cognitive distraction" scale, with its index numbers based on the "hurricane" scale, with "5" indicating what they claim are "catastrophic" safety risks (Figure 4).

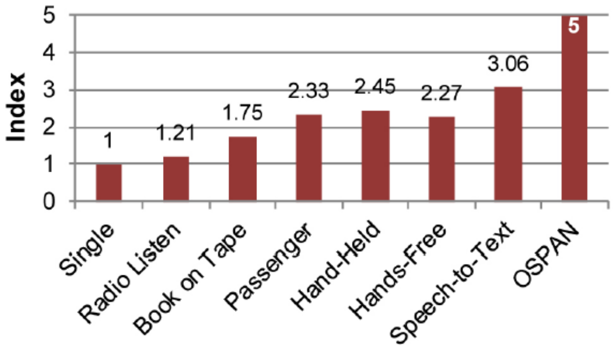


Figure 4. Strayer et al.'s (2013) data on their "cognitive distraction" index scale for 8 speech task conditions. Redrawn from Strayer et al. (2013, Figure 28), and Young (2014a,b, Figure 1, p. 68).

The tasks range from a pure auditory task (Radio Listen), to pure auditory-vocal tasks, with "passenger" "hand-held" and "hands-free" being conversation tasks. In the "operation-word-span" or OSPAN task, originally developed by Turner and Engle (1989), participants are asked to read and verify a simple math problem (such as "Is (4/2)-1=1?), and then read a word after the operation (such as SNOW). After a series of problems and words has been presented, the participants recall the words that followed each operation. The cognitive distraction index score above each bar ranges from 1 for "just driving" with no task, to 5 indicating "catastrophic" increases in safety risk, according to Strayer et al. (2013). The numbers above the bars in Figure 4 are composite standardized scores across multiple variables -- see Strayer et al. (2013, 2015a) for details.

However, Young (2014a,b; 2015a) re-analyzed the original Strayer et al. (2013) data and found that their data do not support their conclusion of increased safety risks from speech tasks, even for the "catastrophic" OSPAN task with an index score of 5. For example, Figure 5 plots Strayer et al.'s own data for their variables of following times (black circles) and brake reaction times (open circles) on the same time scale (*y*-axis) (redrawn from Young, 2015a, Figure 1A) to a lead vehicle in open road driving.

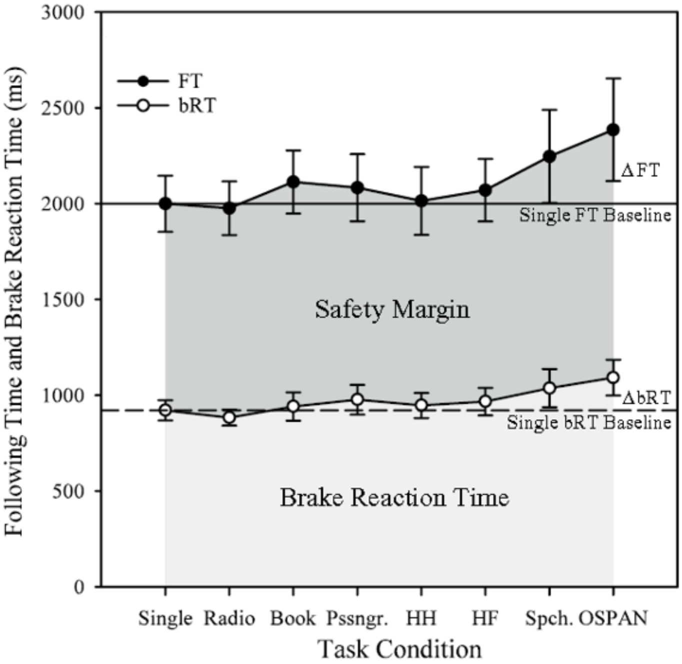


Figure 5. An analysis of the Strayer et al. (2013, 2015a) data showing that drivers compensate for the longer brake reaction time (bRT; open circles) by increasing their following time (FT; black circles), resulting in a larger safety margin (ΔFT) for tasks with higher cognitive distraction scores. Error bars are 95% confidence intervals. Redrawn from Young (2015a, Figure 1A)

The *x*-axis in Figure 5 is composed of the same speech tasks as in Figure 4, again in increasing order of their "cognitive distraction" index score.

The "safety margin" (dark grey area in Figure 5) was defined by Young (2015a) as the following time above and beyond what is needed to make a brake response in time to avoid hitting the lead vehicle, should the lead vehicle suddenly stop. The ΔFT in Figure 5 is the increase in following time compared to the baseline following time with no secondary task (solid horizontal line). The ΔbRT is the increase in brake reaction time compared to baseline (dotted horizontal line). ΔFT increases more than ΔbRT as cognitive demand increases, an example of self-regulation. That is, the drivers slow down sufficiently during an auditory-vocal task such that they have even more time to stop than they would if there were no secondary task.

To illustrate this self-regulation effect more clearly, Figure 6 (top) plots the improvement in the safety margin compared to baseline (the increase in following time or the ΔFT increase above the "single FT Baseline" in Figure 5). Figure 6 (bottom) again plots the Strayer et al. (2013) "cognitive distraction" index from Figure 4, but with error bars.

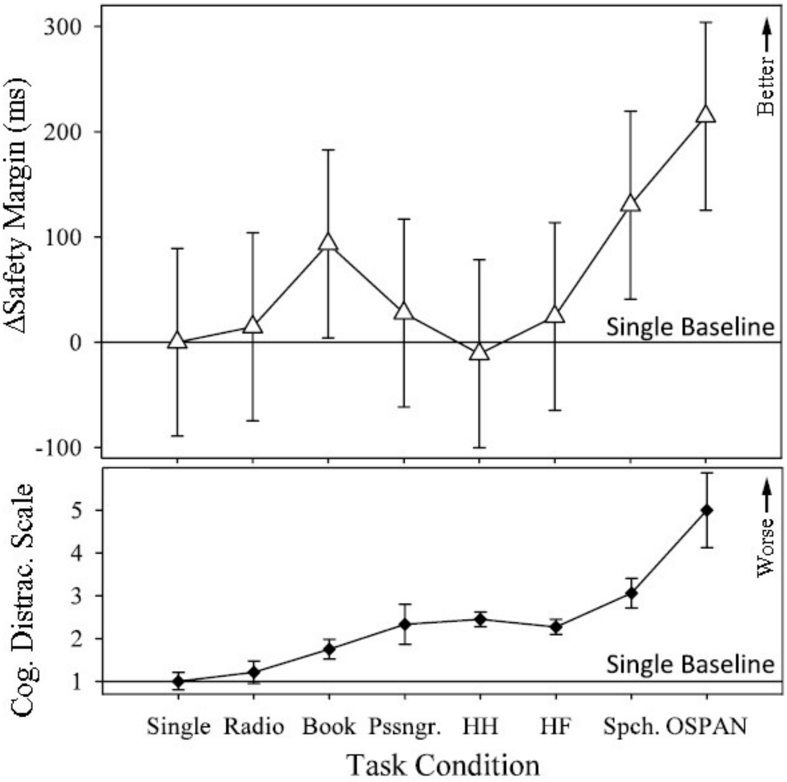


Figure 6. Top. The *y*-axis is the change in safety margin (ΔSafety Margin), which is the difference of ΔFT and ΔbRT in Figure 5 for the task conditions indicated, compared to the Single Baseline (horizontal solid line). Upward indicates a better safety margin. Error bars are pooled 95% confidence intervals. Bottom. Replot of the Strayer et al. cognitive distraction scores in Figure 4. Upward indicates worse "impairment" according to Strayer et al. (2013), compared to the "Single" baseline (horizontal solid line). Error bars are pooled 95% confidence intervals. Redrawn from Young (2015a, Figure 1B)

The correlation between the improved safety margins (top of Figure 6) and the "Cognitive Distraction Scale" (bottom of Figure 6) is $r = 0.90$ ($p = 0.002$). This result indicates that the improvements in safety margin fully compensate for the increase in "cognitive demand" as measured by Strayer and colleagues. The improved safety margin is in fact largest for the tasks with the two highest "cognitive distraction" index scores (text-to-speech and OSPAN).

Young (2014a,b; 2015a) showed that this improvement in the safety margin is the result of driver self-regulation. Strayer et al.'s own data in Figure 5 indicates that drivers slow down as the attention effects of cognitive demand increase. In fact, they slow down so much that the safety margins actually increase more than is needed to fully offset the slight increases in brake reaction time during the speech tasks. These safety margin improvements in Strayer and colleagues' own data, contradict their conclusion of Strayer et al. (2013, 2014, 2015a) that voice tasks cause serious or even catastrophic driver impairments. It follows that their "cognitive distraction" index is not internally valid scientifically, because it does not measure the driver or safety impairments it claims to measure.

In addition to not being internally valid, the Strayer and colleagues' "cognitive distraction" index is also not externally valid, because Strayer et al. assume that it predicts that real-world relative crash risk increases as their "cognitive distraction" index increases. However the opposite occurs for the pure auditory-vocal task of wireless conversation, whether the wireless system is a hands-free phone, hand-held phone, or an integrated wireless device like OnStar

(Young and Schreiner, 2009; Fitch et al., 2013; Klauer et al., 2014) (see Young, 2014b, Figure 5 for summary). Whether a similar downward trend occurs for the Strayer et al. speech-to-text and OSPAN tasks (those with index scores above 3) is not directly determinable, as there is as yet no data on such tasks (or their equivalent) in any real-world or naturalistic driving study to date.

However, a recent proof-of-concept study provides the first evidence that crash risk may actually decrease with increased cognitive demand (Young, 2017a), at least for visual-manual tasks. Visual-manual tasks actually have more cognitive demand than auditory-vocal tasks as shown by Young et al. (2016b) and Appendix A in the current paper. It is hypothesized here that future research will find that auditory-vocal tasks also reduce relative crash risk proportional to their cognitive demand effects in real-world and naturalistic studies. The likely reason is again driver self-regulation (Young, 2014a,b; 2015a), particularly speed decreases.

Later studies by Strayer and colleagues have applied their techniques and "cognitive distraction" index to evaluate what they called "voice" systems in vehicles – both built-in and portable devices; see Cooper et al. (2014) and Strayer et al. (2014; 2015a,b,c). Those studies make the following safety statements without supporting evidence from real-world or naturalistic driving:

> *...voice-based interactions in the vehicle may have unintended consequences that adversely affect traffic safety. (Strayer et al., 2014)*
>
> *...voice-based interactions can be cognitively demanding and ought not to be used indiscriminately while operating a motor vehicle. (Strayer et al., 2015a)*
>
> *Caution is warranted in the use of smartphone voice-based technology in the vehicle because of the high levels of cognitive workload associated with these interactions. (Strayer et al., 2015b)*

Note that these Strayer et al. claims about "voice-based interactions" are actually based on their experiments with mixed-mode tasks with a visual display, rather than pure voice tasks with no visual display. That is, all the "voice" tasks later tested by Strayer and colleagues were actually not "pure" voice tasks because all of their tasks have redundant visual information presented on a display screen (as do the MIT AgeLab studies later discussed in Section 3.4). As shown in Section 3.2 below, mixed-mode tasks have higher attentional effects from cognitive demand, as well as higher physical demand, than do pure auditory-vocal tasks. In short, the assertion by Strayer et al. that "speech" interfaces while driving increase crash risk, much less to "catastrophic " levels, is not supported by their own data and methods, nor by any real-world or naturalistic driving study to date.

For further details on the issues with the Strayer et al. conclusions from their data about increased risk from speech tasks, see Young (2014a,b; 2015a).

3.2 Toyota EntuneTM

Driver performance testing was conducted using a 2014 Toyota Corolla with the navigation option, which included a touchscreen "infotainment" interface (Seaman et al., 2016). This Toyota Entune™ infotainment system includes a touchscreen interface with access to the Toyota Entune™ suite of applications, which, when paired with a smartphone, gave access to the driver apps such as Bing™ for point-of-interest searches and Pandora™ for Internet-based music streaming. The task descriptions and experimental details are given in Seaman et al. (2016).

Figure 7 shows the test set-up.

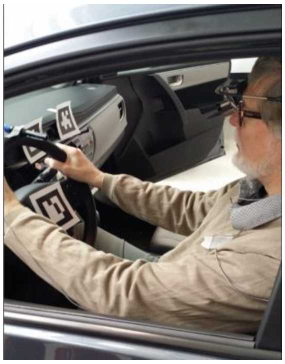


Figure 7. Test set-up for driver performance testing of 2014 Toyota Corolla infotainment system. See Seaman et al. (2016) for further set-up and task details

Glance behavior was recorded and analyzed using the Ergoneers Dikablis eye-tracking system. The pictures taped to the dashboard in Figure 7 are fiducial marks for the eye-tracking system. This system uses a head-mounted infrared eye camera and a forward-facing scene camera to coordinate pupil position with the forward scene. Missing pupil detections and false alarms were corrected by hand. Glance cross-throughs and blinks were removed following the ISO 15007 eye glance standard (ISO 15007-1:2014, 2014; ISO/TS 15007-2:2014, 2014).

Participants performed a surrogate driving test which involved moving the steering wheel to keep a red laser dot in the center of the roadway, which was a video scene projected on a forward screen that was recorded during real driving (Figure 8).

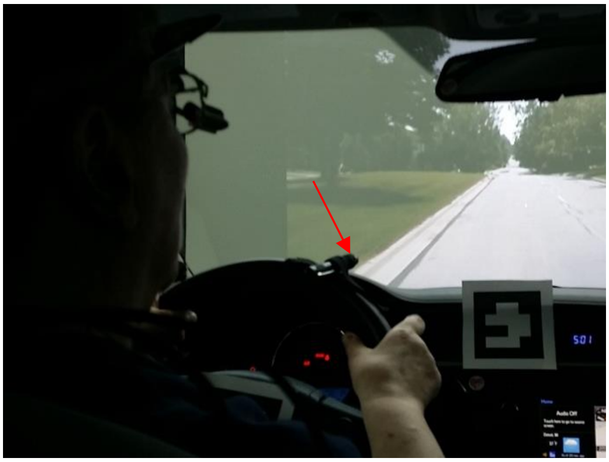


Figure 8. Over the shoulder view of screen with laser (red arrow) at top of steering wheel during driver performance testing of the 2014 Toyota Corolla infotainment system

This test set-up with a video image and some form of detection-response task performance (whether the remote DRT as in Section 2.1 or tactile DRT here) has received extensive validation, and has been shown to well predict driver performance in experiments on a closed or open road (see Section 2.1.1 for citations).

Participants simultaneously performed the tactile detection response task (TDRT) (Hsieh et al., 2015), using the TDRT purchased from TNO in the Netherlands that has been extensively validated in many different sites in various countries in Annex E in the ISO 17488:2016 (2016) detection-response task standard.

There were 15 task conditions:

- A no-task control

- Two pure auditory-vocal tasks, which were an audio presentation / verbal response delayed recall working memory task adapted specifically as a cognitive load reference task for the automotive research environment (Mehler et al., 2011) and presented at what are considered to be low (0-Back) and medium (1-Back) levels of cognitive demand
- Five mixed-mode tasks (all of which had the same start and end points as the visual-manual tasks)
- Seven pure visual-manual tasks

The order of presentation of the 15 secondary task conditions was randomly distributed across the 24 participants, who were balanced for sex and age groups 18-25, 26-35, 36-45, and 46+.

Figure 9 shows the results of a principal components analysis that reduced the 22 mean driver performance variables for each task to two scores for each task in the dimensional space.

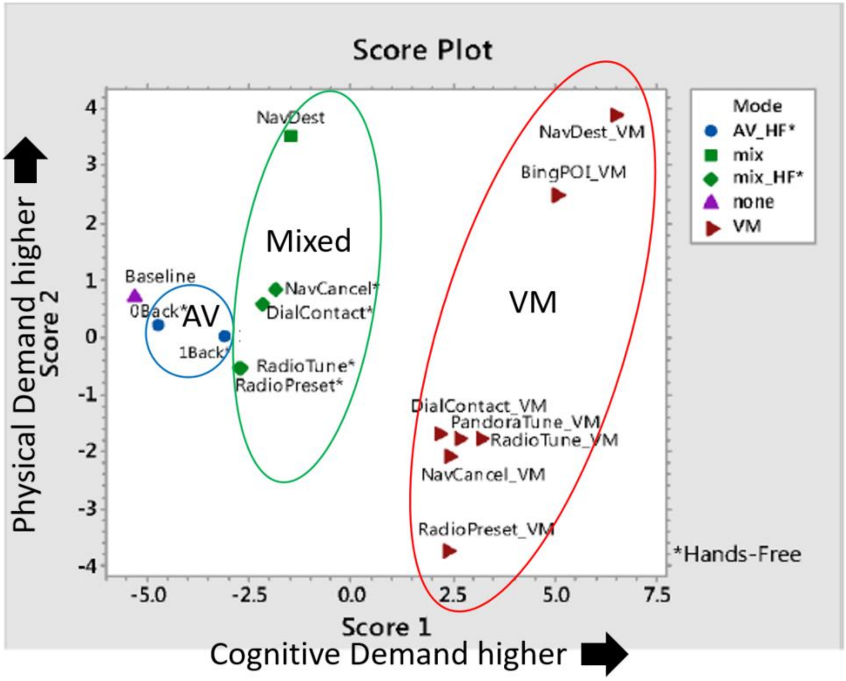


Figure 9. Locations of all 15 task conditions in the study according to their scores on the two major dimensions of driver performance, labeled as "Cognitive Demand" attentional effects on the *x*-axis and "Physical Demand" on the *y*-axis. The colored ellipses show the grouping of the tasks according to the driver interface mode of that task. AV = pure auditory-vocal (blue); mixed = mixed-mode (green); VM = pure visual-manual (red). The baseline (magenta triangle at far left) had 0 secondary tasks. All task conditions included simulated driving, and also responding to a tactile detection response task. Redrawn from Young et al. (2016b, Figure 7)

The major dimension, explaining 60% of total variance in the 22 driver performance metrics, was interpreted as the attentional effects of cognitive demand, because the cluster of metrics associated with it were particularly highly correlated with event detection and response metrics, measured via the Tactile Detection-Response Task (Hsieh et al., 2015; ISO 17488:2016, 2016). The minor dimension, explaining 20% of total variance, was interpreted as physical demand, because the cluster of metrics associated with it were associated with physical actions such as task time, and motor actions such as number of manual steps, number of glances, mouth movements, lung and diaphragm movements for speech, etc. -- see Young et al. (2016b) for details.

Baseline driving with no task (magenta triangle at far left) had the lowest cognitive demand attentional effects as expected. Pure auditory-vocal tasks (filled blue circles within blue oval) had the next highest cognitive demand attentional effect, very near baseline driving with no secondary infotainment task. The mixed-mode tasks (green diamonds within green oval) had intermediate cognitive demand effects. Pure visual-manual tasks (red arrowheads within red oval) had the highest cognitive demand attentional effects, consistent with the results in other studies in Appendix Figure A1 in this review.

The pure speech tasks (0Back and 1Back) have more physical demand than some of the mixed-mode and visual-manual tasks because as mentioned above they require muscle movement of the jaw muscles, and muscle movements of the lungs and diaphragm. In addition, these tasks are conducted for a fixed time of 30 seconds -- they are not self-paced like the other tasks in this study. Hence, they have longer task times than many of the visual-manual tasks below them on the *y*-axis, such as a single button press in the visual-manual mode of the radio preset task (RadioPreset_VM).

Because of the physical demand arising from the muscle movements for speech over their longer task times, Figure 9 indicates that hands-free mixed-mode tasks actually have more physical demand compared to the same task with the same end goal in visual-manual mode. Specifically, Figure 9 shows that RadioPreset, RadioTune, DialContact, and NavCancel in mixed-mode have more physical demand than the same tasks (with the same start and end points) in pure visual-manual mode. The visual screen gives rise to physical demand from eye movements in both modalities. However, the additional mouth, lung, and diaphragm muscle movements for the verbal commands over a longer task time create more physical demand than the hand movements for the visual-manual mode.

3.3 Guidelines for Driver-Vehicle Interaction

Young and Zhang (2015) reviewed current driver behavior metrics and guidelines for visual-manual tasks – including those from the Alliance of Automotive Manufacturers (Alliance, 2006), the National Highway Traffic Safety Administration (NHTSA, 2014), and the Japanese Automotive Manufacturers Association (JAMA, 2004). Guidelines for portable and aftermarket devices were issued by NHTSA (2016). However, guidelines for speech interactions have not yet been published by NHTSA at the time of this writing. Young and Zhang (2015) point out that designing for improved driver performance for any interface modality means minimizing the physical and cognitive demand for drivers. Their design principles are, "Thoughtful reduction of complexity across all layers of user interface design" and "Optimize for the vehicle platform," which are applicable to all interface modalities.

3.4 MIT AgeLab Studies of "Voice-Command" Systems On-Road Studies

The MIT AgeLab conducted numerous on-road field studies to develop empirical data on driver interaction with production-level "voice-command" systems. Their studies to date, performed since Sept. 2012 across 6 vehicle models:

- Study 1 – Focused on the assessment of demands associated with the voice interface, 2010 Lincoln MKS (Reimer et al., 2013).
- Study 2 – The impact of structured vs. unstructured training and replication of Study 1, 2010 Lincoln MKS (Mehler et al., 2014).
- Study 3 – Assessed the impact of an "experienced" user mode, 2010 Lincoln MKS (Reimer et al., 2014).
- Study 4 – Comparisons between embedded vehicle systems and a smart phone, 2013 Chevrolet Equinox and 2013 Volvo XC60 (Mehler et al., 2015a).
- Studies 5-7 – Assessment of generalizability of findings 2014 Chevrolet Impala, 2014 Mercedes CLA and 2015 Toyota Corolla (Mehler et al., 2015b,c,d).
- Study 8 – Voice interface demands under level 1 automated driving, 2014 Chevrolet Impala (Reimer et al., 2015).
- Studies 9-11 – Summaries of many of the above studies (Mehler et al., 2016a,b; Reimer, 2016).

The findings of these extensive MIT AgeLab studies are consistent with the conclusions in Section 3.2 for the Toyota Entune test, and Section 4.1.1 for the Cadillac CUE/iPhone/CarPlay driver interface design analysis. The MIT AgeLab results also concur with the conclusion in

Appendix A1, that substantial cognitive demand effects arise from pure visual-manual interfaces.

However, it is to be noted that significant visual demand is present in the "voice-command" system interfaces tested in these MIT AgeLab studies. Mehler et al. (2015a) recognize this important point, "This study and previous work (e.g. Mehler et al., 2014; Reimer et al., 2013) suggest that the voice interfaces of current embedded systems are highly multi-modal and the full range of potential demands (auditory, vocal, visual, manipulative, cognitive, tactile, etc.) need to be taken into account." Hence, the MIT AgeLab "voice" studies are not what are termed in the current review "pure" voice interfaces; all of the interfaces are what we call "mixed-mode" and the MIT AgeLab researchers call "multi-modal" interfaces. It is therefore not surprising that many of the "voice" tasks they conducted in their field studies exceed the NHTSA (2014) Guidelines glance criteria, which were developed for visual-manual tasks (Reimer, 2016). Note however, that strictly speaking, the NHTSA Guidelines glance criteria were developed for simulator testing, and not on-road testing, and the results for road and simulator testing can be quite different in absolute terms (Young, 2016a; ISO 17488:2016, 2016 Appendix E).

The traditional point of view is that there are separate visual, manual, and cognitive demand channels, with only a small amount of overlap (Figure 10, left). This view is still widely held in the field. For example, a Google search for "visual manual and cognitive distractions while driving" turns up hundreds of references in the last five years, too numerous to cite here. The left-hand graphic in Figure 10 also illustrates that many in the driver interface field classify auditory-vocal tasks as "cognitive," and place visual-manual tasks in the "visual" and "manipulative" classes with little overlap with "cognitive" (Reimer, 2016). However, Appendix A shows that visual-manual tasks actually have far more cognitive demand effects on attention than do auditory-vocal tasks.

For these and other reasons, this traditional point of view needs to be revised, according to the unanimous consensus from a workshop of automotive human factors experts in academia, industry, and government (Foley et al., 2013). The right-hand graphic in Figure 10 (adapted from Reimer, 2016) illustrates the concept in Foley et al. (2013) and Reimer (2016) that cognitive demand effects on attention are central to all forms of demand from all modalities.

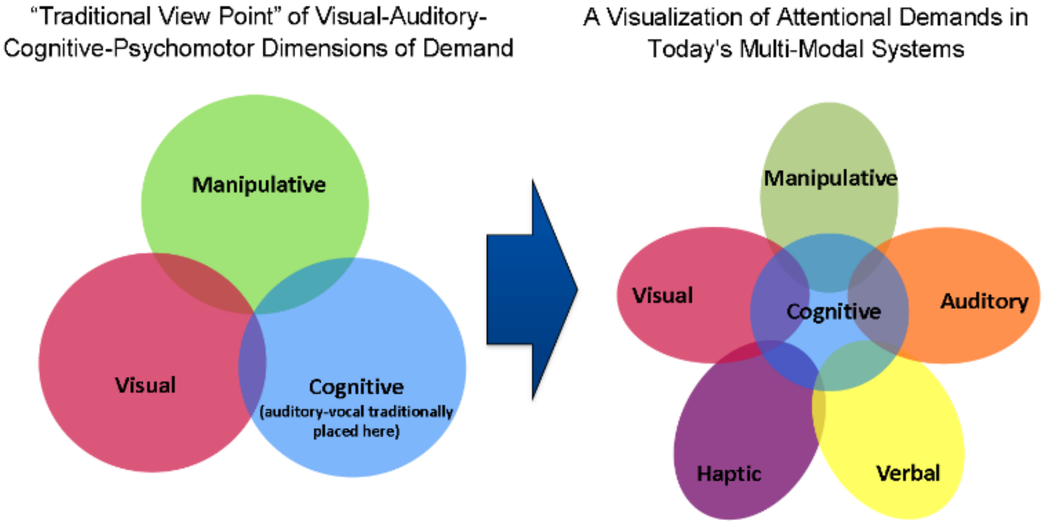

"Traditional View Point" of Visual-Auditory-Cognitive-Psychomotor Dimensions of Demand
Manipulative
Visual
Cognitive
(auditory-vocal traditionally placed here)
A Visualization of Attentional Demands in Today's Multi-Modal Systems
Manipulative
Visual
Cognitive
Auditory
Haptic
Verbal

Figure 10. Left. Traditional concept of the dimensions of demand. Right. Revised concept of the dimensions of demand (Foley et al., 2013). Figure adapted from Reimer (2016)

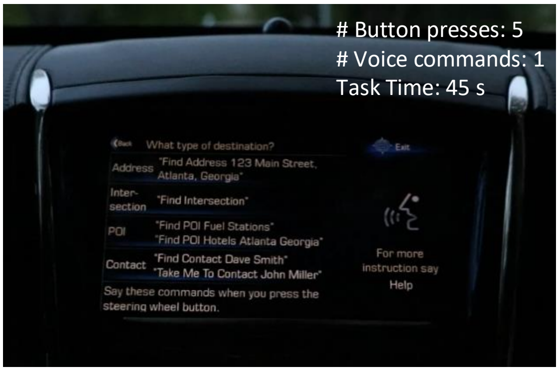
The right-hand graphic in Figure 10 illustrates the concept that no activity that can be done by a human using any modality (haptic, visual, manipulative, auditory, or verbal) that does not involve attentional effects from cognitive demand. Thus, the traditional separation of the demand types into visual, cognitive, and manipulative, with a small overlap of each, is not scientifically accurate, because it implies that most of visual and "manipulative" or manual demand does not require or contribute to cognitive demand. Not only does this traditional viewpoint contradict the data reviewed in this paper, it is not compatible with basic cognitive neuroscience. All human activity, except perhaps for a pure spinal reflex (e.g., from a tap on the patellar tendon), contains an attentional effect of cognitive demand concurrent with the activity. Even with a pure spinal reflex, there would be cognitive activity subsequent to the reflex, assuming the participant were conscious at the time.

4 New Research on Automotive Speech Interfaces

4.1 Cadillac CUE, iPhone, and Apple CarPlay

To check the results and conclusions of the studies reviewed in Sections 2 and 3 of this paper, an informal analytic design evaluation was made of three navigation and communication systems in a recent production vehicle, a 2016 Cadillac ATS. All three systems supported both visual-manual and mixed-mode interfaces for the same tasks, which presents a useful opportunity to compare two interface modes to accomplish the same tasks. The approach was similar to what was done for auditory-vocal and visual-manual modes by Young (1999, 2001a) reviewed in Section 2.1, or for mixed-mode and visual-manual mode by Young and colleagues (Seaman et al., 2016, Young et al., 2016a,b), reviewed in Section 3.2.

Two tasks were tested: entering a navigation destination, and dialing a 10-digit phone number. The design was thus 3 systems x 2 tasks x 2 interface modes. Figures 11-13 show screen shots and results for the three systems during the destination entry task using the mixed-mode interface. Figure 11 is CUE alone, Figure 12 is iPhone alone, and Figure 13 is paired iPhone and CUE (termed *CarPlay*[8]).



# Button presses: 5
# Voice commands: 1
Task Time: 45 s

What type of destination?

Address "Find Address 123 Main Street, Atlanta, Georgia"

Inter-section "Find Intersection"

POI "Find POI Fuel Stations"
"Find POI Hotels Atlanta Georgia"

Contact "Find Contact Dave Smith"
"Take Me To Contact John Miller"

Say these commands when you press the steering wheel button.

For more instruction say Help

Figure 11. CUE alone. The 2016 Cadillac CUE infotainment system by itself during destination entry using a mixed-mode interface, showing redundant visual display on vehicle screen during vocal commands

[8] Retrieved from https://en.wikipedia.org/wiki/CarPlay

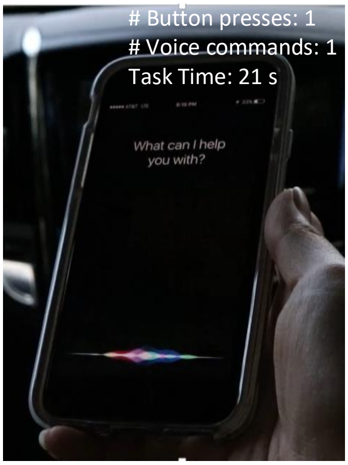
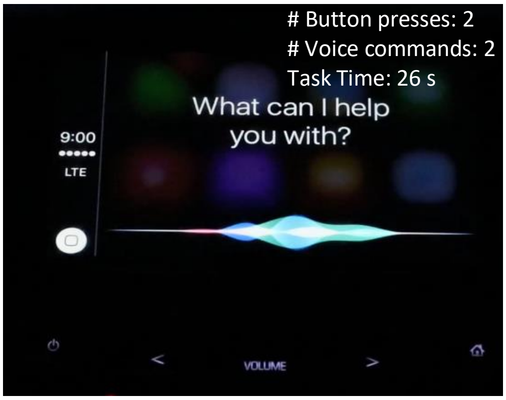


Figure 12. iPhone alone. Destination entry using iPhone with Siri voice system with auditory input to driver from phone speaker, showing redundant visual display on phone screen with moving colored images at bottom



What can I help you with?

Figure 13. CarPlay (paired CUE and iPhone). Destination entry with Siri voice system with auditory input to driver from car speakers, showing redundant visual display on vehicle screen with moving colored images at bottom

This analytical design evaluation simply counted the minimum number of button presses and the minimum number of voice commands required to do the secondary tasks from a fixed start point to a fixed end state. The time to complete the task without error by a highly trained person was also measured. Results are shown in the upper right of Figures 11-13 for the mixed-mode destination entry task, and summarized in Table 1, with the results also for the visual-manual mode for the identical task.

Table 1. Destination Entry task analysis results for visual-manual (VM) and mixed-modes for CUE alone, iPhone alone, and CarPlay (paired CUE and iPhone)

Destination Entry Device	**Mode**	**# Button Presses**	**# Voice Commands**	**Task Time (s)**
CUE alone	VM	24	0	42
	Mixed (Fig. 11)	5	1	45
iPhone alone	VM	35	0	24
	Mixed (Fig. 12)	1	1	21
CarPlay	VM	30	0	33
	Mixed (Fig. 13)	2	2	26

Table 1 shows that the mixed-mode interface required substantially fewer button presses (i.e., number of visual-manual steps) than the visual-manual interface for all three devices. For mixed-mode, the built-in CUE system alone (Figure 11) has 5 manual steps, whereas the iPhone alone (Figure 12) has only 1 manual step, and the iPhone paired with CUE in CarPlay (Figure 13) has only 2 manual steps. That compares with visual-manual which required a minimum of 24, 35, and 30 button presses respectively for destination entry. The fact that there are fewer button presses for mixed mode vs. visual-manual across the three devices indicates that the interface designers of the three systems were successful in swapping out the large number of button presses by using one or two voice commands instead. The findings was similar for the 10-Digit dial task (not shown).

Note that the number of button presses loads onto the physical demand dimension (Young and Angell, 2003; Young, 2012b; Young, 2016a,b). Button presses represent the number of manual steps, which is usually highly correlated with other physical demand variables such as number of glances, eyes-off-road time, perceived subjective workload, task time, etc., in the Dimensional Model of Driver Demand (Young and Angell, 2003; Young, 2012b; Young, 2016a,b). In the current test, the reduced number of visual-manual steps in mixed-mode was achieved without a major increase in the total task time to complete the task using voice commands instead of button presses (last column of Table 1).

Note however that the mixed-mode tasks in the CUE system have an extensive on-screen information display (Figure 11) that is redundant to the audio content. This content would draw driver's eyes off the road and to the screen, as was shown in the MIT AgeLab studies of mixed-mode tasks in Section 3.4. In addition, the iPhone by itself (Figure 12) or paired with CUE (Figure 13) always presents visual text at the beginning of the interactive voice session, "What can I help you with?" that is unnecessary and redundant with the speech output. The iPhone alone or with CarPlay also presents irrelevant moving images (colored moving display at bottom of screen). These moving images are contrary to the final automotive driver distraction guidelines from the Alliance (2006), as well as those from NHTSA (2014, 2016). They are also contrary to one of the NHTSA Guidelines for paired portable devices, "Devices providing dynamic (i.e., moving) non-safety-related visual information should provide a means by which that information cannot be seen by the driver" (NHTSA, 2014, p. 32). This type of dynamic visual information display would be expected to worsen the driver performance (compared to pure auditory-vocal systems) because of unnecessary glances to the display, and must be avoided according to all three sets of guidelines (Alliance, 2006; NHTSA, 2014, 2016).

A shorthand way of making this point is, "If they can see it, they will read it!" which is a conclusion of a recent presentation by Gilbert (2016) on speech interfaces in automobiles. The present recommendations are consistent with this conclusion, as well as his subsequent conclusions that: 1) the optimum human-vehicle interface is truly hands-free and eyes-free, and 2) a mixed-mode (or any interface that includes a visual component) will draw the eyes off the forward roadway. Note that none of the tasks in the visual-manual mode or mixed-mode in Table 1, for either the Cadillac CUE alone, the iPhone alone, or the paired CUE and iPhone (CarPlay), were truly eyes-free because of the redundant and unnecessary visual information (Figures 11-13). Nor were CUE alone and CarPlay hands-free according to the SAE J2972_201403 (2014) definition of hands-free, which permits only 1 button press. Although the iPhone alone (Figure 12) did require only 1 button press, it is held in the hand so it obviously also does not meet the definition of hands-free.

Some may claim that the visual display is glanced at in a mixed-mode interface simply because of a "lack of trust" in pure auditory information, or in voice recognition reliability. Visual glancing is thus seen as confirmatory to the "uncertain" or unfamiliar auditory channel for the driver-vehicle interface. Road tests done by the lead author of the identical auditory-vocal turn-by-turn navigation interface with and without a visual display show that the added visual display does not improve the ability to make the correct turns at the correct time (Young, unpublished). However, in market research clinics, drivers often state that they still "prefer" to have the visual display. Automotive companies therefore need to balance customer preference with automotive safety in such cases.

Some also may claim that such unnecessary glance behavior in mixed-mode displays will diminish over time as voice interfaces gain more ubiquitous acceptance. However, the "orienting reflex” to moving or flashing stimuli in the periphery is just that – a reflex – and there is as yet no data on whether drivers can or will successfully inhibit this reflex with exposure over time.

Note that the Apple Guidelines document for developers that was current at the time of writing this review places the emphasis on how to program applications for the iPhone that keep the users “fully engaged with the screen,” specifically in order “to improve revenue for developers and advertisers.” There are no statements in the Apple Guidelines that discuss designing for use while driving, where the goal is exactly the opposite – to keep the driver’s eyes on the road, hands on the wheel, and mind on the drive, not on the screens or controls of the iPhone. Apple CarPlay may reduce the information content from the standalone iPhone, which may make it more suitable for use while driving. However, it is an open question how many iPhone users actually pair their iPhones with CarPlay, even assuming their vehicle model even supports CarPlay. For drivers with older vehicle models, or drivers who are unable or unwilling to use CarPlay in newer vehicle models, they will use the iPhone in standalone mode.

4.2 Android Auto

Google’s Android Auto system is a mechanism for “projecting” the Android operating system smartphone application display onto the built-in screen of an automotive infotainment system (Figure 14).

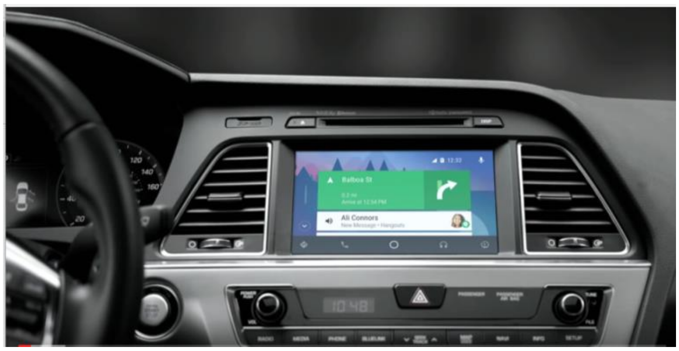


Figure 14. Screen shot of Android Auto

Google’s public comments about Android Auto make clear that Google recognizes that applications for Android Auto must be designed differently for use while driving than not driving. A 39-minute video explaining Google’s approach was publicly released on May 18, 2016.[9] In the video, Google specifically states that “Designing for the phone is not the same thing as designing for driving.” In addition, they point out that designing for driving is, “…very different than designing for the phone, where you design for full engagement.” Google states in the video that the company recognizes that if digital connection is not done the right way, it could be a problem for what they call “safe” driving. They add, “Driving and keeping your eyes on the road is Task #1, and dealing with the phone should be Task #2.”

The “Designing for Driving” principles that Google advocates in the video9 are:

- Biased towards action
- Simple, predictable and non-hierarchical
- Enable Voice in All Apps
- Say “OK Google” to eliminate need to press voice button
- Built-in typography, touch targets, contrast for driving (vibration, dynamic lighting issues during day, day vs. night lighting)
- Knowable and familiar

[9] Retrieved from https://www.youtube.com/watch?v=JiwiTDXTO78

- Naturally integrated into car
- Adapted to technologies in car
- Differentiation for competition
- Multi-dimensional problem – different countries, driving habits, apps & services, many cars
- Common Platform
- Design process: Responsible to many
- User experience vs. human performance
- Humans not as capable in car while driving
- Iterative research and design process
- Test everything!

Google has established the "Google Research Lab for Android Auto" driving simulator laboratory and are actively engaging in testing Google products in their simulator, as the video demonstrates.9

The Google video9 demonstrates at least an initial attempt at addressing driver performance issues with smartphones with an appropriate human factors design and testing process. Again, as with the iPhone, for drivers with older vehicle models, or drivers who are unable or unwilling to use Android Auto in newer vehicle models, they will use their Android smartphone in standalone mode.

4.3 Dimensional Model of Driver Demand Extended to All Modalities

The Extended Dimensional Model of Driver Demand (Young et al., 2017b) is hypothesized to further simplify the six demand types on the right side of Figure 10, into only two demand types, physical and cognitive demand. This hypothesis is illustrated in Figure 15.

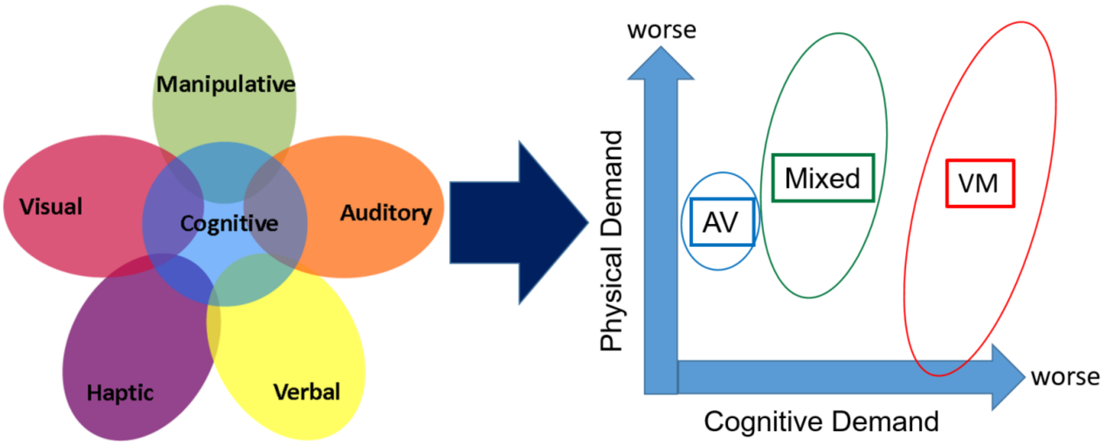


Figure 15. Left. The six demand types from Reimer (2016), from Figure 10. Right. Reduction of the six demand types to only two demand types, according to the Dimensional Model in Figure 9. Task modes: AV = pure audio-vocal mode, Mixed = mixed-mode, VM = pure visual-manual mode.

The hypothesis is that the Extended Dimensional Model simplifies all the driver performance metrics for all possible demand modes and tasks on the left of Figure 15 (Reimer, 2016), into only the two axes or dimensions in Figure 15, right (Young et al., 2017b). The Model therefore includes all six demand types and modes that can have an effect on driver performance in a highly simplified manner. The lead author has interpreted these axes or dimensions as *physical* and *cognitive* demand – see Young et al. (2016b) and Figure 9 in the current review.

See Young et al. (2016a) for further explanation of the Dimensional Model for visual-manual tasks, and Young and Angell (2003) and Young (2012b) for earlier versions. Young et al. (2016b) extended the Dimensional Model from visual and manual modalities to include auditory and verbal modalities in the physical and cognitive demand dimensions, as illustrated in Figure 9 and the right side of Figure 15. In sum, we advance the new hypothesis that the Dimensional Model will encompass all modalities (including haptic) into a single integrated cognitive and physical demand 2-Dimensional model. This is a new hypothesis which requires further investigation and testing.

5 Review Limitations

As mentioned in Section 1, this review is not meant to be exhaustive of all the dozens of studies that have been conducted on automotive speech technologies. It instead reflects studies that the authors have direct experience with, or otherwise believe are particularly important for speech interface development for automobiles, from their viewpoint.

However, these results cannot necessarily be generalized to all secondary tasks that can be performed in a vehicle. Some secondary tasks may have better performance in pure auditory-vocal mode, while others may be better in pure visual-manual mode, and still others in mixed-mode. Conceivably, some secondary tasks may be even impossible in pure auditory-vocal mode (for example getting an overview over two possible routes to a destination). Nor would the critical primary driving tasks of steering and braking be suitable for an auditory-vocal interface, because steering adjustments with voice tend to take more time than manual adjustments of the steering wheel, and likewise for brake pedal presses. Thus, the scope of this review's conclusion about the driver performance benefits of the auditory-vocal mode is limited to those categories of tasks that have been tested to date. It is not possible to say on the basis of these results, that a pure auditory-vocal mode would have better driver performance than a pure visual-manual mode for all possible tasks that can be performed in a vehicle.

6 Future Research Recommendations

6.1 Emotional Tone, Music Effects on Driver Performance

Several studies have been conducted on the effect on driver performance of the emotional tone of speech output by the vehicle, or listening to music of different types. However, this topic is not as well researched as other sections of this review, so it is presented here under the section for future research recommendations, as it would benefit from future investigation.

6.1.1 Unmatched Emotional Tone Effects

Studies have shown that some emotional tones can improve driver performance without monitoring the driver's emotions and matching the emotional tone to the driver's emotions (what we call *unmatched* tone).

For example, and contrary to common expectation, an angry speech tone by the vehicle voice system improves (i.e., decreases) driver response times to a visual detection-response task compared to neutral speech (Seaman et al., 2008, 2010; Hsieh et al., 2009, 2010a,b). This effect occurs both in the laboratory (Seaman et al., 2008) and on the open road (Hsieh et al., 2009, 2010a). These behavioral improvements were also observed directly in the underlying brain areas that gave rise to the behavior using functional magnetic resonance imaging in a brain imaging study with surrogate driving with the identical voice system and detection response task (Hsieh et al., 2010b). Other experiments used confirmed this result using event-related potential recordings from the scalp in laboratory and open road studies, with the identical voice system, detection response task, and event-related potential recording equipment (Hsieh et al., 2009, 2010a).

Similar findings using behavioral tests and event-related potentials from the scalp were found simply by using words with positive or negative emotional valence, during simulated driving and non-driving conditions (Chan and Singhal, 2015). Negative emotion valence words slowed driving speed, compared to positive and neutral. Both positive and negative word valences caused improved driving performance compared to neutral, probably because of an arousal effect.

6.1.2 Music Effects

A related line of research is the effect of listening to different types of music while driving. Music effects are outside the scope of this review of speech effects on driving performance, but a number of music studies have shown strong effects on driving performance, both positive and negative depending upon the music volume, tempo, and content. The book by Brodsky (2015) contains a comprehensive review.

6.1.3 Matched Emotional Tone Effects?

The late Clifford Nass and colleagues (Nass and Brave, 2005; Nass et al., 2005) conducted a provocative and controversial simulator study on the effect of emotional content in automotive speech on driving performance. The study suggests that speech user interfaces may reduce crashes by matching the simulated emotions of the system to those of the driver. The experiment was performed in a driving simulator with 20 female and 20 male participants. All participants drove on the same simulated course for approximately fifteen minutes. Half of the participants had an induced "happy" emotional state at the time of the experiment, and the other half an induced "upset" emotional state. These emotional states were induced by showing half of the participants (randomly selected) seven minutes of very happy video clips, and showing the other half seven minutes of very upsetting video clips. A questionnaire confirmed that individuals had the intended emotional state before they entered the driving simulator. The study then tested the effects of automated verbal remarks at thirty-six separate points along the driving course. For half of the happy participants and half of the upset participants (randomly selected), the voice was energetic and upbeat; for the other half, the voice was calm and subdued. The simulator automatically recorded the number of crashes that drivers were involved in during their driving time in the simulator.

Nass and colleagues found that matching the tone of the remarks to the induced emotions of the driver had a substantial effect on the number of crashes of the simulated cars. Figure 16 plots their results.

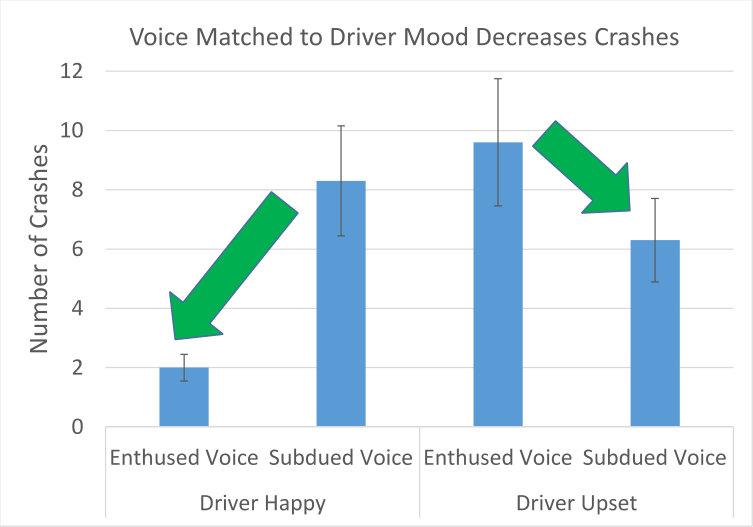


Figure 16. Plot of the data from Nass and Brave (2005) and Nass et al. (2005)

Figure 16 illustrates that drivers who interacted with voices that matched their own emotional state (enthused voice for happy drivers and subdued voice for upset drivers) had less than half as many (simulated) crashes on average as drivers who interacted with mismatched voices. The green arrows in Figure 16 illustrate the reduction in crashes from matching the voice emotion to the driver emotion. When combined with vehicle systems that could infer the emotional state of the driver, such as facial emotion recognition using a non-intrusive video camera, this data suggests that automatically adjusting the emotional cues conveyed by an automotive speech driver-vehicle interface might improve the crash reduction benefits of voice systems.

However, this study has been criticized by speech recognition experts as an interesting study but in a specific context that bears replication and further research before attempting to implement in production vehicles. In particular, the experiment lacked an important control: a moderately-toned control voice – neither upbeat nor subdued. Other limitations to its generalizability are the artificially-induced emotions, which may or may not correspond to the emotions drivers might experience in real-world driving. There were also a non-naturalistic overabundance of opportunities to crash, and an unrealistically small number of only two emotional states. Once moving beyond two emotions, the likelihood of a voice tone mismatch would only increase just because of mathematical probability. For this matched-emotion approach to be successful in production, there would need to be almost perfect determination of the driver's emotional state because, if the system were wrong, the use of the incorrect voice would, if these results held in the real world, *increase* the likelihood of crashing (or at least diminish the driver's performance). It would be difficult for automotive companies to determine how often the correctly chosen voice tone improved driver performance and crash avoidance, except during a naturalistic study which instruments vehicles to measure crashes and driver behavior. Such a study would also need to record audio as well as video, which is not common practice in many naturalistic driving studies. A legal hurdle to overcome would be whether an "incorrect" voice tone would conceivably increase automaker liability in crash-related lawsuits?

In addition, without even a replication during on-road driving (such as described in Sections 2.2, 2.3, 3.1 and 3.4), it simply is not clear whether the findings of Nass and colleagues will generalize from the simulator to having strong effects on voice systems in use during on-road driving. It is also not clear that these emotional tone variables have sufficient influence to improve upon the qualities of an already usable voice system and tone — one that is already efficient, effective, and pleasant. For example, Couper, Singer, and Tourangeau (2004) studied the influence of male and female artificial voices on more than 1000 respondents to an interactive voice-recognition survey on sensitive topics, including testing for similarity-attraction. Measurement variables included respondents' reactions to the different voices and abandoned call rates. They found no statistically significant results related to the apparent gender of the voices. In particular, there were no significant gender-by-voice by gender-by-respondent interactions. "Why such strong effects of humanizing cues are produced in laboratory studies but not in the field is an issue for further investigation" (Couper et al., 2004, p. 567).

Before implementing speech tone variation in production vehicles, future research studies should therefore be done that more fully investigate the effects of emotional speech tones on driver performance, particularly in on-road studies such as those of Hsieh et al. (2009, 2010a).

6.2 Relative Crash Risk from Speech Interfaces?

Many of the simulator experiments in the previous sections were not validated in on-road driving experiments, much less real-world driving or naturalistic driving conditions. Driver self-regulation in more realistic driving conditions tends to offset the negative effects of wireless speech conversation on driving performance observed in laboratory experiments (Young, 2014a,b; 2015a,b; 2017b) (see Section 3.1).

A critical safety question for future research on speech interfaces is the relative crash risk of auditory-vocal interfaces when measured in real-world or naturalistic driving studies.

In the Dimensional Model solution for tasks of different modalities, the attentional effect of cognitive demand is the dominant axis, explaining 60% of the total variance in driver performance. Thus, the effects of cognitive demand effects form the dominant *x*-axis in Figure 9 and the right-side graphic of Figure 15. This axis represents event detection and response in the human brain. The specific neural location for the response time effects of the detection–response task in a driving-like scenario during speech tasks is known to be localized in the right superior parietal lobe in the human brain (Bowyer et al., 2009). This region of the brain is known to be part of the orienting attention network, which selects information from sensory input (Posner and Fan, 2008). In other words, the ability to orient and respond to events is synonymous with the attentional effects of cognitive demand. In turn, event detection and response is obviously critical to driving safety (Angell, 2008, 2010). The direct effect of reduced event detection ability from cognitive demand should therefore be an increase in relative crash risk (i.e., poorer driving safety). Auditory-vocal interfaces, since they increase response times and depress activity in this part of the brain, would thus be expected to lead to increased crash risk vs. not doing any secondary task at all. A second hypothesis that follows from this line of analysis, is that auditory-vocal tasks would still be predicted to be better for driving safety compared to visual-manual or mixed-mode interfaces because of their relatively lower attentional effects of cognitive demand (Figure 9 and Appendix A).

However, the first hypothesis (that performing auditory-vocal tasks increase crash risk compared to no task) is not supported by real-world and naturalistic driving data. Personal cell phone conversation while driving, whether by hand-held, hands-free, or embedded phone, does not increase and may actually reduce relative crash risk (i.e., improves driving safety) compared to no cell phone conversation (Young and Schreiner, 2009; Young, 2012a, 2013, 2014c, 2017c). The first paper to show that wireless conversation while driving has a low absolute crash risk was Young (2001b). This low crash risk from wireless conversation (despite the likely increases in response time known from experimental studies) was attributed to driver self-regulation (Young, 2014a,b; 2015a; 2017a), such as slowing down (Young, 2014b).

The cognitive demand of visual-manual tasks has a similar reduction in crash effect. Indeed, the cognitive demand effects of visual-manual tasks have an almost perfect negative correlation with relative crash risk. This counterintuitive result was demonstrated in a recent proof-of-concept study that correlated the cognitive demand effects of tasks on a track, with crashes in a naturalistic driving study while performing similar tasks (Young, 2017a). That is, for visual-manual tasks, the stronger the effects of cognitive demand on attention, the more the odds of a crash decreased. Again, the likely reason is driver self-regulation.

There are insufficient auditory-vocal tasks on the road today (other than cell phone conversation or voice interfaces on smartphones) to estimate the relative risk as a function of the attentional effects of cognitive demand for auditory-vocal tasks, as was done for visual-manual tasks, in a naturalistic driving or real-world crash study (Young, 2017a). Such a study will require future research, if and when speech interfaces become more widely used in automobiles, and naturalistic driving studies record audio as well as video.

7 Summary and Recommendations

The studies reviewed in this paper compared different modalities for completing a task across many studies: pure auditory-vocal vs. pure visual-manual modes vs. mixed-mode. The results across studies were consistent in finding that:

- A pure auditory-vocal interface reduced the attentional effects of cognitive demand vs. a pure visual-manual mode interface (Sections 2.1, 2.2, 2.3; Appendix A1)
- A pure auditory-vocal interface had better safety margins as one measure of "cognitive demand" increased (Section 3.1, Figures 5 and 6), likely because of driver self-regulation
- A pure auditory-vocal interface reduced the attentional effects of cognitive demand vs. a mixed-mode interface (Section 3.2, Figure 9)
- A mixed-mode interface reduced the attentional effects of cognitive demand vs. a pure visual-manual mode (Section 3.2, Figure 9; Section 3.4; Section 4, Table 1).

These findings provide empirical evidence for the hypotheses of Gilbert (2016) that: (1) the optimum human-vehicle interface should be truly hands-free and eyes-free, and (2) a mixed-mode (or any interface that includes a visual component) will draw the eyes off the forward roadway.

We recommend that future speech interface products should sensibly limit redundant head-down visual displays in order to improve driver performance. In fact, we recommend speech-only driver-vehicle interfaces (with no head-down visual display) to the extent possible. If well-implemented, such pure speech interfaces can have good customer acceptance in market research clinics, and potentially in the marketplace. If widely implemented and accepted by the driving public, speech interfaces are predicted to give rise to improved driving safety over visual-manual interfaces because of their comparatively lower attentional effects and reduced off-road glances compared to mixed-mode and pure visual-manual interfaces. It is recommended that future naturalistic driving studies be designed that can investigate whether pure speech interfaces actually reduce relative crash risk as predicted by cell phone conversation naturalistic driving studies, compared to mixed-mode or visual-manual interfaces.

8 Conclusions

The results of this review indicate that using a pure speech interface for secondary tasks while driving is predicted by a number of experimental studies to give rise to better driver performance in real-world driving than using a pure visual-manual interface. Likewise, a pure speech interface is predicted to give rise to better driver performance in real-world driving than a mixed-mode interface with a head-down visual display that provides redundant information to the speech interface. The visual component of mixed-mode and visual-manual interfaces increases not only cognitive demand effects on attention, but also reduces eyes-off-road time compared to a pure voice interface.

Acknowledgments

We thank Jim Kortge, Katja Kircher, and three anonymous reviewers for useful comments, and Adelle Victor for assistance with evaluating the Cadillac CUE, iPhone and Apple CarPlay user interfaces. Portions of this paper were presented by the first author as an invited Keynote Address at the SpeechTek 2016 conference in Washington, D.C. (Young, 2016b). The authors received no funding for writing this paper from any outside source.

Definitions/Abbreviations

AV Auditory-vocal interface mode

CAMP-DWM Crash Avoidance Metrics Partnership Driver Workload Metrics project

DRT Detection Response Task. See ISO 17488:2016 (2016), and RDRT and TDRT definitions.

HUD Head-up display, projected onto windshield

J2365 An SAE best practice on how to calculate, measure, and predict task time for visual-manual tasks. Auditory-vocal tasks are specifically excluded from the scope. See SAE J2365_201607 (2016).

J2972 Automotive industry definition of hands-free operation for wireless communication. See SAE J2972_201403 (2014).

J2988 Automotive industry guidelines for driver-vehicle speech interfaces. See SAE J2988_201506 (2015).

mixed-mode Auditory-vocal and visual-manual modes in the same driver task. A task that is performed with hearing and speaking but has just a visual interface (with no button presses required) is also a mixed-mode task.

mule vehicle An existing production vehicle that has been converted to a pre-production concept vehicle.

naturalistic driving The driver’s own vehicle is equipped with video cameras that record the driver’s behavior. Other instruments also record the vehicle’s behavior, in real time, while a vehicle is driven in everyday fashion over a prolonged period, from months to several years. Naturalistic driving is a subset of real-world driving.

primary driving tasks The operational tasks of driving *per se* which are critical to driving: namely, steering, pressing and releasing the accelerator, braking, and detecting and responding with an appropriate steering or braking maneuver to objects and events in the roadway. In vehicles with manual transmissions, primary tasks would also include pressing and releasing the clutch pedal and operating the gearshift lever. Other tasks that are critical to the driving task are also sometimes defined as primary, including speedometer checks, mirror/blind spot checks, and activating wipers/headlights. (See “secondary tasks” definition.)

pure A *pure* auditory-vocal task means that the task has just auditory and vocal modes alone, without visual displays or manual input. Likewise, a *pure* visual-manual task means the task has only just visual and manual interface modes, without auditory output from the vehicle system, or voice-recognition by the vehicle system. The term “pure” is intended to distinguish these auditory-vocal and visual-manual modes from mixed-mode (see definition of *mixed-mode*).

real-world driving Driving a vehicle in an everyday manner, without experimental instructions or special equipment such as video cameras installed in the vehicle that are not part of the original vehicle equipment (see naturalistic driving)

RDRT Remote Detection-Response Task. A task in which a small light appears at a distance away from the driver, and the driver must detect and respond to it as rapidly as possible, often with the press of a button attached to the finger. See ISO 17488:2016 (2016).

SAE Society for Automotive Engineers

secondary tasks Tasks performed in a vehicle by a driver that are not related to the primary driving tasks (i.e., that are not critical for driving). Some non-critical vehicle tasks are defined as secondary tasks in some studies, but as primary tasks in others, such as radio adjustments, seatbelt adjustments, window adjustments, vent adjustments, visor and mirror adjustments, and setting cruise control.

TDRT Tactile Detection-Response Task. A task in which a small vibrator on the shoulder activates, and a driver must detect and respond to it as rapidly as possible, often with the press of a button attached to the finger. See ISO 17488:2016 (2016).

UI User Interface

VM Visual-Manual interface mode

Appendix A. Lower Attentional Effects of Cognitive Demand for Auditory-Vocal vs. Visual-Manual Tasks

Figure A1 shows that drivers have better (i.e., faster) responses to visual events during pure auditory-vocal tasks (A,B), than they do during pure visual-manual tasks (C).

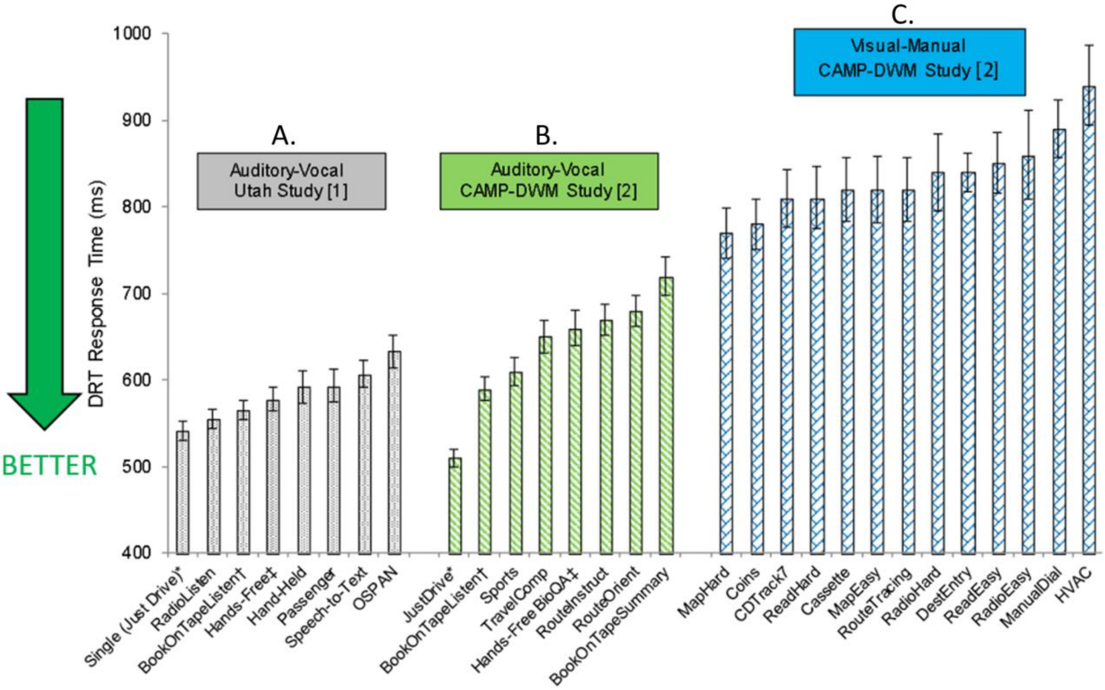


Figure A1. Comparison of Remote Detection-Response Task (DRT) response time to the onset of a small light away from the driver during pure auditory-vocal vs. pure visual-manual tasks in two simulator studies. Note: smaller bars indicate between driver performance (faster response times). A: Response times during pure auditory-vocal tasks (Strayer et al., 2013). B: Response times during pure auditory-vocal tasks in the CAMP-DWM study (Angell et al., 2006a,b). C: Response times during pure visual-manual tasks in the CAMP-DWM study (Angell et al., 2006a,b). Redrawn from Young (2014b, Figure 6).

The detection-response task response time is an effective measure of the attentional effects of cognitive demand (ISO 17488:2016, 2016). Therefore, Figure 1A demonstrates that pure visual-manual tasks tend to have a worse attentional effect from cognitive demand than do pure auditory-vocal tasks. This result holds even for the pure auditory-vocal task with the longest (worst) response time in the Strayer et al. (2013) study, the OSPAN task (rightmost bar in A). Likewise, the worst auditory-vocal task in the CAMP-DWM study (BookOnTapeSummary, rightmost bar in B), still had a shorter bar (lower response time and cognitive demand effect) vs. the least cognitively-demanding visual-manual task in the CAMP-DWM study (MapHard) (C).

This better performance for pure auditory-vocal tasks may in part arise because of the greater eyes-off-road time for pure visual-manual tasks compared to pure auditory-vocal tasks. Eyes away from the forward event light will increase response times. However, visual-manual tasks also have longer response times than auditory-vocal tasks even when using the tactile detection response task (Figure 9 in main body), a haptic stimulus that is not influenced by eye movements. This better performance for auditory-vocal tasks is further confirmed by the ISO detection-response task validation studies (ISO 17488:2016, 2016, Annex E), which also used the tactile detection response task. The ISO studies compared a pure visual-manual task (Surrogate Reference task) with pure auditory-vocal tasks (1Back and 0Back), and found that the response times were longer for the visual-manual than auditory-vocal tasks, whether measured during surrogate driving, simulator driving, or on-road driving.

References

Alliance. (2006). Statement of principles, criteria and verification procedures on driver-interactions with advanced in-vehicle information and communication systems, June 26, 2006 version. Washington, DC: Alliance of Automobile Manufacturers Driver Focus - Telematics Working Group. Retrieved from http://www.autoalliance.org/index.cfm?objectid=D6819130-B985-11E1-9E4C000C296BA163

Angell, L. (2008, June 2). Conceptualizing effects of secondary task demands on event detection during driving: Surrogate methods & issues. Presentation at the *Driver Performance Metrics Workshop*, San Antonio, Texas. Retrieved from http://drivingassessment.uiowa.edu/drivingmetrics/P_Conceptualizing%20Event%20Response%20Linda.pdf

Angell, L. (2010). Conceptualizing effects of secondary task demands on event detection during driving: Surrogate methods and issues. In G. L. Rupp (Ed.), *Performance metrics for assessing driver distraction: The quest for improved road safety* (Chapter 3, pp. 42-71). Warrendale, PA USA: SAE International.

Angell, L. S., Young, R. A., Hankey, J. M., & Dingus, T. A. (2002, May). An evaluation of alternative methods for assessing driver workload in the early development of in-vehicle information systems. Paper from the *Society of Automotive Engineers Government/Industry Meeting*, Washington, DC. Retrieved from http://www.researchgate.net/publication/228557232

Angell, L., Auflick, J., Austria, P., Kochhar, D., Tijerina, L., Biever, W., Diptiman, T., Hogsett, J., & Kiger, S. (2006a). Driver workload metrics task 2 final report. (DOT HS 810 635). National Highway Traffic Safety Administration, Crash Avoidance Metrics Partnership (CAMP). Retrieved from http://www.nhtsa.gov/DOT/NHTSA/NRD/Multimedia/PDFs/Crash%20Avoidance/Driver%20Distraction/Driver%20Workload%20Metrics%20Final%20Report.pdf

Angell, L., Auflick, J., Austria, P., Biever, W., Diptiman, T., Hogsett, J., Kiger, S., Kochhar, D., & Tijerina, L. (2006b). Driver workload metrics project, task 2 final report, appendices. National Highway Traffic Safety Administration, Crash Avoidance Metrics Partnership (CAMP). Retrieved from http://www.nhtsa.gov/DOT/NHTSA/NRD/Multimedia/PDFs/Crash%20Avoidance/2006/Driver%20Workload%20Metrics_appendices.pdf

Bowyer, S., Hsieh, L., Moran, J., Young, R., Manoharan, A., Liao, C.-C., Malladi, K., Yu, Y.-J., Chiang, Y.-R., & Tepley, N. (2009). Conversation effects on neural mechanisms underlying reaction time to visual events while viewing a driving scene using MEG. *Brain Research*, *1251*, 151-161. Retrieved from http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2741688/, doi:10.1016/j.brainres.2008.10.001

Brodsky, W. (2015). *Driving with music: Cognitive-behavioural implications.* Ashgate Publishing, Ltd.

Chan, M., & Singhal, A. (2015). Emotion matters: Implications for distracted driving. *Safety Science*, *72*, 302-309. Retrieved from http://www.sciencedirect.com/science/article/pii/S0925753514002331, doi:https://doi.org/10.1016/j.ssci.2014.10.002

Cooper, J. M., Ingebretsen, H., & Strayer, D. L. (2014). Mental workload of common voice-based vehicle interactions across six different vehicle systems. Washington, DC: AAA Foundation for Traffic Safety. Retrieved from https://www.aaafoundation.org/sites/default/files/Cog%20Distraction%20Phase%20IIA%20FINAL%20FTS%20FORMAT.pdf

Couper, M. P., Singer, E., & Tourangeau, R. (2004). Does voice matter? An interactive voice response (IVR) experiment. *Journal of Official Statistics*, *20*(3), 551-570. Retrieved from https://www.researchgate.net/profile/Roger_Tourangeau/publication/313646390

Fitch, G. M., Soccolich, S. A., Guo, F., McClafferty, J., Fang, Y., Olson, R. L., Perez, M. A., Hanowski, R. J., Hankey, J. M., & Dingus, T. A. (2013). The impact of hand-held and hands-free cell phone use on driving performance and safety-critical event risk final report. Washington, D.C.: NHTSA. Retrieved from http://www.nhtsa.gov/DOT/NHTSA/NVS/Crash%20Avoidance/Technical%20Publications/2013/811757.pdf

Foley, J., Young, R., Angell, L., & Domeyer, J. (2013), June 17-20). Towards operationalizing driver distraction. Paper from the *7th International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design*, Bolton Landing, New York. Retrieved from http://drivingassessment.uiowa.edu/sites/default/files/DA2013/Papers/010_Foley_0.pdf

Gilbert, J. (2016). An examination of speech-enabled technologies in the car. Presentation at the *SpeechTEK 2016*, Washington, D.C. Abstract retrieved from http://www.speechtek.com/2016/Wednesday.aspx

Hsieh, L., Seaman, S., Sullivan, J., Bowyer, S., Moran, J., Angell, L., Young, R. (2009). Effects of emotional speech tone of cell phone conversations on driving: ERP, lab and on-road driving studies. Presentation at the *Cognitive Neuroscience Society*, San Francisco, CA. Retrieved from https://www.researchgate.net/publication/259621371

Hsieh, L., Seaman, S., & Young, R. A. (2010a). Effect of emotional speech tone on driving from lab to road: fMRI and ERP studies. Paper from the *AutomotiveUI 2010: 2nd International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, Carnegie Mellon University, Pittsburgh, PA, U.S.A. Retrieved from https://www.auto-ui.org/10/proceedings/p22.pdf

Hsieh, L., Seaman, S., Jiang, Q., Bowyer, S., Moran, J., & Young, R. (2010b, April 19). Neural basis of emotional modulation of simulated driving performance: An fMRI multitasking study. Presentation at the *Cognitive Neuroscience Society*, Montreal, Quebec, Canada. Retrieved from http://www.academia.edu/20167873/Neural_basis_of_emotional_modulation_on_of_simulated_driving_performance-_An_fMRI_multitasking_study-abstract2_poster-ok

Hsieh, L., Seaman, S., & Young, R. A. (2015). A surrogate test for cognitive demand: Tactile detection response task (TDRT). SAE Technical Paper 2015-01-1385. Retrieved from https://www.researchgate.net/profile/Richard_Young9/publication/275353622, doi:10.4271/2015-01-1385

ISO 15007-1:2014(en). (2014). Road vehicles — measurement of driver visual behaviour with respect to transport information and control systems — part 1: Definitions and parameters. International Standards Organization. Retrieved from https://www.iso.org/obp/ui/#iso:std:iso:15007:-1:ed-2:v1:en

ISO 17488:2016. (2016). Road vehicles -- Transport information and control systems -- Detection-Response Task (DRT) for assessing attentional effects of cognitive load in driving. 76 pages. Retrieved from http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=59887

ISO/TS 15007-2:2014. (2014). Road vehicles -- measurement of driver visual behaviour with respect to transport information and control systems -- part 2: Equipment and procedures. International Standards Organization. Retrieved from http://www.iso.org/iso/catalogue_detail.htm?csnumber=56622

JAMA. (2004). Guideline for in-vehicle display systems — version 3.0. Retrieved from http://www.jama-english.jp/release/release/2005/jama_guidelines_v30_en.pdf

Klauer, S. G., Guo, F., Simons-Morton, B. G., Ouimet, M. C., Lee, S. E., & Dingus, T. A. (2014). Distracted driving and risk of road crashes among novice and experienced drivers. *New England Journal of Medicine*, *370*(1), 54-59. Retrieved from http://www.nejm.org/doi/full/10.1056/NEJMsa1204142

Lutz, B. (2011). *Car guys vs. bean counters: The battle for the soul of American business*. New York, NY: Portfolio/Penguin

Mehler, B., Reimer, B., & Dusek, J. (2011). MIT AgeLab delayed digit recall task [n-back]. Retrieved from https://www.researchgate.net/profile/Bruce_Mehler/publication/230729111

Mehler, B., Reimer, B., Dobres, J., McAnulty, H., Mehler, A., & Coughlin, J. (2014). Further evaluation of the effects of a production level "voice-command" interface on driver behavior: Replication and a consideration of the significance of training method. Technical Report 2014-2. Cambridge, MA: MIT AgeLab. Retrieved from https://www.researchgate.net/publication/265594707

Mehler, B., Kidd, D., Reimer, B., Reagan, I., Dobres, J., & McCartt, A. (2015a). Multi-modal assessment of on-road demand of voice and manual phone calling and voice navigation entry across two embedded vehicle systems. *Ergonomics*, 1-24. Retrieved from http://jdobr.es/pdf/Mehler-etal-2015-IIHS-Embedded-Interfaces.pdf

Mehler, B., Reimer, B., Dobres, J., & Coughlin, J. F. (2015b, November). Assessing the demands of voice based in-vehicle interfaces - Phase II experiment 1 - 2014 Chevrolet Impala. Cambridge, MA USA: Massachusetts Institute of Technology. Retrieved from https://www.researchgate.net/publication/301198003

Mehler, B., Reimer, B., McAnulty, H., Dobres, J., Lee, J., & Coughlin, J. F. (2015c, November). Assessing the demands of voice based in-vehicle interfaces - Phase II experiment 2 - 2014 Mercedes CLA. Cambridge, MA USA: Massachusetts Institute of Technology. Retrieved from https://www.researchgate.net/publication/301198159

Mehler, B., Reimer, B., Dobres, J., & Coughlin, J. F. (2015d, November). Assessing the demands of voice based in-vehicle interfaces - Phase II experiment 3 - 2015 Toyota Corolla. Cambridge, MA USA: Massachusetts Institute of Technology. Retrieved from http://agelab.mit.edu/files/Publications/Mehler_etal_MIT_AgeLab_Techical%20Report_2015-14_Corolla.pdf

Mehler, B., Reimer, B., Dobres, J., Foley, J., & Ebe, K. (2016a). Additional findings on the multi-modal demands of "voice-command" interfaces. SAE Technical Paper 2016-01-1428. Retrieved from http://jdobr.es/pdf/Mehler-etal-2016-SAE.pdf, doi:10.4271/2016-01-1428

Mehler, B., Kidd, D., Reimer, B., Reagan, I., Dobres, J., & McCartt, A. (2016b). Multi-modal assessment of on-road demand of voice and manual phone calling and voice navigation entry across two embedded vehicle systems. *Ergonomics*, *59*(3), 344-367. Retrieved from https://pdfs.semanticscholar.org/203d/439fb637484a76a981aeb7d25e991bc20c78.pdf, doi:10.1080/00140139.2015.1081412

Microsoft (1998). Microsoft announces Auto PC, PC companion powered by Windows CE 2.0, Retrieved from http://news.microsoft.com/1998/01/08/microsoft-announces-auto-pc-pc-companion-powered-by-windows-ce-2-0/

Nass, C. I., & Brave, S. (2005). *Wired for speech: How voice activates and advances the human-computer relationship*: MIT press, Cambridge, MA.

Nass, C., Jonsson, I.-M., Harris, H., Reaves, B., Endo, J., Brave, S., & Takayama, L. (2005). Improving automotive safety by pairing driver emotion and car voice emotion. Paper from the *CHI'05 Extended Abstracts on Human Factors in Computing Systems*, Portland, OR, USA. Retrieved from http://www.leilatakayama.org/downloads/Takayama.CarEmotion_CHI2005_prepress.pdf

NHTSA. (2014, September 26). *Guidelines for reducing visual-manual driver distraction during interactions with integrated, in-vehicle, electronic devices version 1.01* (docket no. NHTSA–2010–0053, ID NHTSA-2014-0088-0002). Washington, DC: National Highway Traffic Safety Administration. Retrieved from http://www.regulations.gov/contentStreamer?documentId=NHTSA-2014-0088-0002&attachmentNumber=1&disposition=attachment&contentType=pdf

NHTSA. (2016, December 5). *Visual-manual NHTSA driver distraction guidelines for portable and aftermarket devices* (docket no. NHTSA–2013–0137*)*. Washington, DC: National Highway Traffic Safety Administration. Retrieved from https://www.gpo.gov/fdsys/pkg/FR-2016-12-05/pdf/2016-29051.pdf

Posner, M. I., & Fan, J. (2008). Attention as an organ system. In J. R. Pomerantz (Ed.), *Topics in integrative neuroscience: From cells to cognition* (pp. 31–61). Cambridge, UK: Cambridge University Press. Retrieved from https://www.researchgate.net/publication/248544743, doi:10.1017/CBO9780511541681.005

Reimer, B. (2016, September 27-28). Evolving HMI assessment toward a multi-modal vision in an increasingly automated driving experience. Presentation at the *SID 2016 23rd Annual Symposium on Vehicle Displays: Vehicle Displays and and Interfaces*, Livonia, MI USA. Retrieved from pp. 264-304 in http://vehicledisplay.org/VehicleDisplay_2016.pdf

Reimer, B., Mehler, B., Dobres, J., & Coughlin, J. (2013). The effects of a production level “voice-command” interface on driver behavior: Summary findings on reported workload, physiology, visual attention, and driving performance. MIT AgeLab White Paper. Retrieved from http://agelab.mit.edu/files/MIT_AgeLab_White_Paper_2013-18A_(Voice_Interfaces).pdf

Reimer, B., Mehler, B., Dobres, J., & Coughlin, J. (2015, December 10). Phase II experiment 4 – An exploratory study of driver behavior with and without assistive cruise control (ACC). Cambridge, MA USA: Massachusetts Institute of Technology Technical Report 2015-15. Retrieved from http://jdobr.es/pdf/Reimer-etal-2015-ACC.pdf

Reimer, B., Mehler, B., Dobres, J., McAnulty, H., Mehler, A., Munger, D., & Rumpold, A. (2014). Effects of an 'expert mode' voice command system on task performance, glance behavior & driver physiology. Paper from the *6th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, Seattle, WA USA. Retrieved from http://agelab.mit.edu/files/Reimer_et_al_2014_Auto-UI_Voice_Expert_Mode.pdf

SAE J2365_201607. (2016). Calculation and measurement of the time to complete in-vehicle navigation and route guidance tasks. Warrendale, PA: SAE International. Retrieved from http://standards.sae.org/j2365_201607/

SAE J2972_201403. (2014). Definition of road vehicle hands-free operation of a person-to-person wireless communication system or device. Society of Automotive Engineers. Retrieved from http://standards.sae.org/j2972_201403/

SAE J2988_201506. (2015). Guidelines for speech input and audible output in a driver vehicle interface. SAE International. Retrieved from http://standards.sae.org/j2988_201506/.

Schreiner, C. (2006). The effect of phone interface and dialing method on simulated driving performance and user preference. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *50*(22), 2359-2363. Retrieved from http://journals.sagepub.com/doi/abs/10.1177/154193120605002202, doi:10.1177/154193120605002202

Schreiner, C., Blanco, M., & Hankey, J. (2004a). Investigation of driving behavior changes associated with manual and voice-activated phone-dialing in a real-world environment. Paper from the *At the Crossroads: Integrating Mobility Safety and Security*. ITS America 2004, 14th Annual Meeting and Exposition. Retrieved from https://trid.trb.org/view.aspx?id=704229

Schreiner, C., Blanco, M., & Hankey, J. M. (2004b, September 20-24). Investigating the effect of performing voice recognition tasks on the detection of forward and peripheral events. Paper from the *HFES 2004: Proceedings of the Human Factors and Ergonomics Society 48th Annual Meeting*, New Orleans, Louisiana. Retrieved from https://www.researchgate.net/publication/274042766, doi: 10.1177/154193120404801932

Seaman, S., Hsieh, L., & Young, R. (2008). The effect of emotional conversation on visual detection during simulated driving: A behavioral study. Presentation at the *Cognitive Neuroscience Society*, San Francisco, CA. Retrieved from https://www.researchgate.net/publication/259602791

Seaman, S., Hsieh, L., Wu, L., & Young, R. A. (2010). Neural basis of emotional modulation while multitasking: ERP analysis. Presentation at the Cognitive Neuroscience Society, Montreal, Quebec, Canada. Retrieved from https://www.researchgate.net/publication/259621403

Seaman, S., Hsieh, L., & Young, R. (2016). Driver demand: Eye glance measures. SAE Technical Paper 2016-01-1421. Retrieved from https://www.researchgate.net/publication/301258124, doi:10.4271/2016-01-1421

Strayer, D. L., Turrill, J., Cooper, J. M., Coleman, J. R., Medeiros-Ward, N., & Biondi, F. (2015a). Assessing cognitive distraction in the automobile. Human Factors, 57(8), 1300-1324. Retrieved from https://sites.tufts.edu/appliedcognition/files/2015/10/Assessing-cognitive-distraction-in-the-automobile.pdf, doi:10.1177/0018720815575149

Strayer, D. L., Cooper, J. M., Turrill, J., Coleman, J. R., & Hopman, R. J. (2015b). Measuring cognitive distraction in the automobile III: A comparison of ten 2015 in-vehicle information systems. Washington, DC USA: AAA Foundation for Traffic Safety. Retrieved from https://www.aaafoundation.org/sites/default/files/strayerIII_FINALREPORT.pdf

Strayer, D. L., Cooper, J. M., Turrill, J., Coleman, J. R., & Hopman, R. J. (2015c). The smartphone and the driver's cognitive workload: A comparison of Apple, Google, and Microsoft's intelligent personal assistants. Retrieved from https://www.aaafoundation.org/sites/default/files/strayerIIIa_FINALREPORT.pdf

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. Retrieved from https://www.aaafoundation.org/sites/default/files/MeasuringCognitiveDistractions.pdf

Strayer, D. L., Turrill, J., Coleman, J. R., Ortiz, E. V., & Cooper, J. M. (2014). Measuring cognitive distraction in the automobile II: Assessing in-vehicle voice-based interactive technologies. Retrieved from https://www.aaafoundation.org/sites/default/files/Cog%20Distraction%20Phase%202%20FINAL%20FTS%20FORMAT_0.pdf

Turner, M. L., & Engle, R. W. (1989). Is working memory capacity task dependent? *Journal of memory and language*, *28*(2), 127-154. Retrieved from https://www.researchgate.net/profile/Randall_Engle/publication/259397774

Young, R. A. (1999, May 24). Communication interface and the older driver. Invited address, *Gerontology Conference*, Ann Arbor, MI, USA. Retrieved from https://www.researchgate.net/publication/259602922

Young, R. A. (2001a, June). Communication interface, vision and the older driver. Invited address from the *Eyes on Design Conference*, Auburn Hills, MI, USA. Retrieved from https://www.researchgate.net/publication/259602742

Young, R. A. (2001b, August 14-17). Association between embedded cellular phone calls and vehicle crashes involving airbag deployment. *Proceedings of Driving Assessment 2001: The First International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design*, Aspen, CO. Retrieved from http://drivingassessment.uiowa.edu/DA2001/81_young-richard.pdf

Young, R. A. (2012a). Cell phone use and crash risk: Evidence for positive bias. *Epidemiology*, *23*(1), 116-118. Retrieved from https://www.researchgate.net/publication/51797905, doi:10.1097/EDE.0b013e31823b5efc

Young, R. A. (2012b). Event detection: The second dimension of driver performance for visual-manual tasks. *SAE Int. J. Passeng. Cars - Electron. Electr. Syst.*, *5*(1), 297-316. Retrieved from https://www.researchgate.net/publication/259602632, doi:10.4271/2012-01-0964

Young, R. A. (2013, September 4-6). Cell phone conversation and automobile crashes: Relative risk is near 1, not 4. Paper from the *Third International Conference on Driver Distraction and Inattention*, Gothenburg, Sweden. Retrieved from http://document.chalmers.se/download?docid=cfd54630-edad-4476-b145-bd46fc08d9b7

Young, R. A. (2014a). Self-regulation reduces crash risk from the attentional effects of cognitive load from auditory-vocal tasks. Paper from the *Transportation Research Board 93rd Annual*

Meeting, No. 14-1010. Retrieved from https://www.researchgate.net/publication/272153134

Young, R. A. (2014b). Self-regulation minimizes crash risk from attentional effects of cognitive load during auditory-vocal tasks. *SAE Int. J. Trans. Safety*, *2*(1), 67-85. Retrieved from https://www.researchgate.net/publication/261472492, doi:10.4271/2014-01-0448

Young, R. A. (2014c). An unbiased estimate of the relative crash risk of cell phone conversation while driving an automobile. *SAE Int. J. Trans. Safety*, *2*(1), 46-66. Retrieved from https://www.researchgate.net/profile/Richard_Young9/publication/261472300, doi:10.4271/2014-01-0446

Young, R. A. (2015a). Driver compensation: Impairment or improvement? *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *57*(8), 1334-1338. Retrieved from https://www.researchgate.net/publication/283536341, doi:10.1177/0018720815585053

Young, R. A. (2015b). Cell phone conversation and relative crash risk. In Y. Zheng (Ed.), *Encyclopedia of mobile phone behavior* (Chapter 102, pp. 1274-1306). Hershey, PA, USA: IGI Global. Retrieved from https://www.researchgate.net/publication/313595382, doi:10.4018/978-1-4666-8239-9.ch102g

Young, R. A. (2015c). Revised odds ratio estimates of secondary tasks: A re-analysis of the 100-car naturalistic driving study data. SAE Technical Paper 2015-01-1387. Retrieved from https://www.researchgate.net/publication/275353775, doi:10.4271/2015-01-1387

Young, R. A. (2016a). Evaluation of the total eyes-off-road time glance criterion in the NHTSA visual-manual guidelines. *Transportation Research Record: Journal of the Transportation Research Board* (*2602*), 1-9. Retrieved from https://www.researchgate.net/publication/313748900, doi:10.3141/2602-01

Young, R. A. (2016b, May 23-35). Driven to distraction? Speech technologies in the automobile. Invited keynote address at *SpeechTek 2016*, Washington DC. Retrieved from http://conferences.infotoday.com/documents/249/0900_Young.pdf

Young, R. A. (2017a). Predicting relative crash risk from the attentional effects of the cognitive demand of visual-manual secondary tasks. SAE Technical Paper 0148-7191. Retrieved from https://www.researchgate.net/profile/Richard_Young9/publication/315859034, doi:10.4271/2017-01-1384

Young, R. A. (2017b). Removing biases from crash odds ratio estimates of secondary tasks: A new analysis of the SHRP 2 naturalistic driving study data. SAE Technical Paper 2017-01-1380. Retrieved from https://www.researchgate.net/publication/315866780, doi:10.4271/2017-01-1380

Young, R. A. (2017c, in press). Cell phone conversation and relative crash risk: Extension. In Y. Zheng (Ed.), *Encyclopedia of mobile phone behavior*. Hershey, PA, USA: IGI Global.

Young, R. A., & Angell, L. S. (2003, July 21-24). The dimensions of driver performance during secondary manual tasks. Paper from the *Driving Assessment 2003: The Second International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design*, Park City, Utah. Retrieved from http://drivingassessment.uiowa.edu/DA2003/pdf/25_Youngformat.pdf

Young, R. A., & Schreiner, C. (2009). Real-world personal conversations using a hands-free embedded wireless device while driving: Effect on airbag-deployment crash rates. *Risk Analysis*, *29*(2), 187-204. Retrieved from https://www.researchgate.net/publication/23464587, doi:10.1111/j.1539-6924.2008.01146.x

Young, R. A., & Zhang, J. (2015). Safe interaction for drivers: Driver behavior metrics and design implications. SAE Technical Paper 2015-01-1384. Retrieved from https://www.researchgate.net/publication/275353879, doi:10.13140/RG.2.1.2150.8641

Young, R. A., Aryal, B., Muresan, M., Ding, X., Oja, S., & Simpson, S. N. (2005). Road-to-lab: Validation of the static load test for predicting on-road driving performance while using advanced in-vehicle information and communication devices. *Proceedings of the Third*

International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design. Retrieved from http://drivingassessment.uiowa.edu/DA2005/PDF/35_DickYoungformat.pdf

Young, R. A., Seaman, S., & Hsieh, L. (2016a). The dimensional model of driver demand: Visual-manual tasks. *SAE Intl. J. Trans. Safety*, *4*(1), 33-71. Retrieved from https://www.researchgate.net/profile/Richard_Young9/publication/301272700, doi:10.4271/2016-01-1423

Young, R. A., Hsieh, L., & Seaman, S. (2016b). The dimensional model of driver demand: Extension to auditory-vocal and mixed-mode tasks. *SAE Int. J. Trans. Safety*, *4*(1), 72-106. Retrieved from https://www.researchgate.net/profile/Richard_Young9/publication/301272732, doi:10.4271/2016-01-1427

About the Authors

Richard Young
Specializes in driver performance and safety effects of cognitive and physical demand. Until 2009, GM Global Driver Workload Lead. Until 2016, Research Professor with joint appointments in School of Medicine and College of Engineering, Wayne State University. Consults for automotive companies, governments, and courts, through his company Driving Safety Consulting.

Jing Zhang
Transforms ideas into prototypes and products where automotive meets digital. Her research innovation and design direction shaped the Chevrolet Volt HMI, OnStar RemoteLink app, and FCA Uconnect apps. Jing runs her design consultancy AutoSimpler, teaches design at Lawrence Technological University, and leads a human interface studio at Fiat Chrysler Automobiles.