Fonds de recherche
Société et culture
Québec

Rapport de recherche
PROGRAMME ACTIONS CONCERTÉES

BOURSES DE RECHERCHE POSTDOCTORALES

Nouvelles technologies et conduite automobile : bénéfices et risques à la conduite pour différents groupe d'âge de conducteurs

Chercheur postdoctoral

Martin Lavallière, Ph.D., Massachusetts Institute of Technology, AgeLab, Cambridge (MA), États-Unis

Nom des cochercheurs ou des chercheurs collaborateurs

Bryan Reimer, Massachusetts Institute of Technology, AgeLab, Cambridge (MA), États-Unis
Bruce Mehler, Massachusetts Institute of Technology, AgeLab, Cambridge (MA), États-Unis
Jonathan Dobres, Massachusetts Institute of Technology, AgeLab, Cambridge (MA), États-Unis
Joseph F. Coughlin, Massachusetts Institute of Technology, AgeLab, Cambridge (MA), États-Unis

Numéro du projet de recherche

2013-3O-171570

Titre de l'action concertée

Programme de recherche en sécurité routière FRQSC, SAAQ, FRQS

Partenaire(s) de l'Action concertée

La Société de l'assurance automobile du Québec (SAAQ),
le Fonds de recherche du Québec - Santé (FRQS)
et le Fonds de recherche du Québec - Société et culture (FRQSC)

Saving lives
through research
and education

Evaluating Technologies Relevant to the Enhancement of Driver Safety

2014

AAA Foundation for Traffic Safety

607 14th Street, NW, Suite 201 | Washington, DC 20005 | AAAFoundation.org | 202-638-5944

Title

Evaluating Technologies Relevant to the Enhancement of Driver Safety *(August 2014)*

Authors

Bruce Mehler, Bryan Reimer, Martin Lavallière, Jonathan Dobres, and Joseph F. Coughlin
Massachusetts Institute of Technology AgeLab

Acknowledgments

A number of people have contributed significant time, thought, and experience during the development of this project, ranging from early discussions about factors that might be important in better understanding the real-world effectiveness of emerging safety technologies, to the identification of relevant information, and the consideration of how available data might be scaled and presented in a manner that is both understandable and which constructively contributes to a dialog around the further enhancement of driving safety. These individuals represent a cross-section of perspectives ranging from industry experts, academic researchers, safety advocates, current and former government employees with expertise in the area, and so on. Many of these contributions were provided on personal time. While all input was taken seriously and considered in the development of this report, it was ultimately taken as advisory. Final decisions about what was included in this report, the language used, and conclusions expressed are the responsibility of the authors and do not necessarily represent the opinions or formal positions of any of the contributors mentioned, the organizations they are affiliated with, the AAA Foundation for Traffic Safety, or individual AAA Clubs. With the aforementioned in mind, the authors would like to extend their sincere appreciation to everyone who commented on or otherwise contributed to this undertaking and, in particular, to AAAFTS for initiating and supporting the development of this work.

About the Sponsor

AAA Foundation for Traffic Safety
607 14th Street, NW, Suite 201
Washington, DC 20005 202-638-5944
www.AAAFoundation.org

Founded in 1947, the AAA Foundation in Washington, D.C. is a not-for-profit, publicly supported charitable research and education organization dedicated to saving lives by preventing traffic crashes and reducing injuries when crashes occur. Funding for this report was provided by voluntary contributions from AAA/CAA and their affiliated motor clubs, from individual members, from AAA-affiliated insurance companies, as well as from other organizations or sources.

Table of Contents

Executive Overview

The AAA Foundation for Traffic Safety tasked the MIT AgeLab with developing a data-driven system for rating the effectiveness of new in-vehicle technologies intended to improve driver safety. Such a system was envisioned as having the potential to educate and guide consumers towards more confident and strategic purchasing decisions, ideally encouraging adoption of technologies showing demonstrated safety benefit. Further, an evaluation of the status and extent of existing data was seen as a way of identifying research gaps in the present state of knowledge about these safety systems. It should be made clear that the focus of the project based on the mandate given to the MIT AgeLab was on given technologies as a class, not on a rating review of individual vehicle model implementations. Development of the rating system and identification of data was undertaken in consultation with identified academic, industrial, consumer, NGO, and governmental experts as well as with representatives of a majority of the major automotive manufacturers. Almost universal endorsement of the importance of this undertaking was voiced.

A top level rating structure has been developed that independently considers safety benefit potential and objectively demonstrated benefit. The latter values are often found to be lower than theoretical expectations. Factors that may be relevant to understanding why such differences appear have been identified as part of the overall project. The emphasis on ratings based on observed benefit for actual drivers under real-world conditions is a key aspect of why this system complements, rather than competes with, ratings developed by IIHS and NCAP which focus largely on controlled test track evaluations of engineered capability. In addition, the rating structure assesses benefit relative to overall crash, injury, and fatality rates – and in relation to the specific scenario / crash event type that a given technology is intended to address. This allows consumers to consider a technology relevant to their particular driving needs.

A total of seven technologies have been reviewed – two reference technologies (Electronic Stability Control and Adaptive Cruise Control) and five emergent safety technologies (Adaptive Headlights, Back-Up Cameras, Forward Collision Warning, Forward Collision Mitigation, and Lane Departure Warning). A major finding of the project has been that only relatively limited data is available upon which to objectively rate the real-world performance of most of these safety systems. A number of experts and industry representatives expressed some surprise at both the divergence between theoretical and observed benefit and the relative scarcity of data upon which to make objective assessments, while others were quite aware of these issues and the need for the development of objective data on real-world performance. This undertaking appears to have already succeeded in one of its goals by stimulating substantive constructive discussion and engagement within the research and industry based safety communities.

The report begins with a review of the original project objectives and an overview of key activities undertaken during the course of the project. Selected observations on the development of the proposed rating system follow, including a brief discussion of the evolving view of rating factors and concepts of scaling that were considered as the project developed. The issue of rating a technology class in contrast to rating specific vehicle implementations is summarized. Key concepts in the proposed rating system are then presented. In particular, the concepts of projected vs. demonstrated benefit and overall safety benefit vs. scenario specific benefit are discussed. A proposed approach to scaling ratings is presented. Detailed reviews of available data on each of the technologies were developed as a core component of the project and used to identify values to rate the technologies. The benefit values extracted from the reviews to assign ratings to each technology class are described. To appreciate the full context from which the ratings are drawn, readers are encouraged to see the more extensive technology review summaries provided in Appendix C. These reviews not only consider the issue of evaluating technology effectiveness, but also address topics including consumer awareness and trust, mobility significance, technology penetration, frequency of use, training and educational issues associated with the technologies, behavioral adaptation, demands placed upon the driver, vehicle type considerations, limitations and failure conditions, and differences between technology implementations. Several approaches to summarizing the ratings in a matrix format are presented. These are intended to be conceptual in nature rather than necessarily representing exact design format recommendations.

The report then presents a number of observations arising out of the work on the project. As already noted, a major point of discussion is the issue of the relatively limited body of objective data that is available to advise not only the consumer, but also the wider public and automotive industry on how these emerging safety technologies are actually performing. Also emphasized is the point that what is presented in this report is a proposed rating system; it is intended to encourage discussion and consideration of important issues related to the better understanding of safety technologies. The extent to which the approach and values proposed here are further developed and updated to keep pace with rapidly evolving technologies and increased scientific evaluation of their performance is seen as an open question. An extended consideration of limitations and points to be kept in mind follows. Comments and critiques from industry reviewers and various Advisory Panel members are recognized and integrated in this important section.

Project Objectives

This project was initiated in response to an RFP issued by the AAA Foundation for Traffic Safety (AAAFTS) entitled *Effectiveness and Efficacy of Technologies to Reduce Older Driver Crashes*. Following discussions with the Foundation's R&D Committee, it was decided to expand the scope of the project to consider drivers across the lifespan.

The stated purpose of the RFP was to support research to identify and develop objective measures that could be used to construct a rating system to compare and contrast the effectiveness of a wide range of relatively new in-vehicle technologies that are relevant to enhancing driver safety. At the initiation of the project, we stated that we believed that this undertaking has the potential to provide a useful tool to aid consumers in making informed decisions about the purchase and use of safety systems. It was also seen as an opportunity to stimulate discussion and possible action within the safety research and manufacturer communities relative to developing a more comprehensive approach to thinking about safety and methods for objectively evaluating new technologies. Our experience interacting with numerous safety research, automotive manufacturer, and technology supplier professionals over the course of the project suggests that it has already had some impact on the latter front.

Selected Observations on the Development of a Rating System

Of Factors and Scaling: An Evolving View

A major focus in the RFP was on the identification of factors relevant to understanding the effectiveness of various technologies in terms of enhancing safety. Emphasis was also placed on developing ratings derived primarily from unbiased objective measures. The vision called for reporting the ratings in the form of a matrix showing how a technology was rated, and envisioned the matrix as a tool that could be used to inform consumers on the value and usefulness of various in-vehicle safety technologies. The RFP included the following list as illustrative examples of what might be considered as factors:

- Estimates of the number of crashes/injuries/fatalities reduced based on previously published research or new estimates developed;
- Cognitive workload measures;
- Driving performance as measured by a simulator and/or instrumented car;
- Efficacy of the technology; and
- Self-reported use of the technology (e.g. deactivation or use of "bypass" feature).

Consequently, we defined task 1 of the project to be a comprehensive review of potential factors for inclusion in a rating system. The review was to take into consideration advantages and disadvantages of each factor based upon a synthesis of the scientific literature, public availability of objective data, feasibility of primary data collection where needed, potential for consumer understanding and impact factors on driver safety. A great deal of thought went into developing an initial listing of potential factors and these were reviewed and refined over the course of the project, as highlighted in the outline of project activities, through extensive discussions with selected experts, industry representatives, and with the Advisory Panel. Aspects of this process can be seen in various materials reproduced in the Appendices.

It gradually became clear over the course of the project that the identification of factors that are relevant to understanding the effectiveness and overall safety benefit associated with a given technology and the development of a consumer oriented method of objectively rating the safety benefit of a given technology are, in fact, best considered as somewhat distinct undertakings. Factors such as technical limitations of the basic technology, usability / understandability of the technology, behavioral adaptation, added cognitive demand / confusion associated with technology implementations, etc. are all important factors to be taken into consideration at the technical design, integration, and evaluation levels by professionals involved in assessing given technologies. However, the construction of a rating matrix or matrices that take into consideration all such potentially meaningful factors rapidly become quite complex; this complexity would be difficult for a typical consumer to follow. Moreover, it has become quite evident in our search of data sources that many of the cells in such a matrix would be found to be empty due to a lack of appropriate data. Where data do exist, it is often difficult to integrate values / findings across cells since available data is often specific to different implementations of a given technology type – thus not being directly comparable. These, and other considerations, make it functionally impractical, if not impossible, to develop objectively defined, well scaled, ratings of overall safety benefit directly from such an underlying factor matrix. On the other hand, the typical consumer is likely to be less concerned with what the underlying factors are that might explain why a given technology offers more or less benefit than they are with being provided with some understanding of what the relative level of benefit appears to be.

This has led us to view the project as developing two different types of evaluation. The "top level" rating system is intended to be consumer oriented and focuses on available objective data on the safety benefit of a given technology. At this level, the operative question is "what is the relative safety benefit?" as opposed to focusing on what might account for the apparent level of benefit. The majority of the main body of this report is directed at detailing a proposed method for making such a top level rating of safety benefit that can be presented to the consumer.

The second type of evaluation considers the underlying factors that may account for the apparent level of safety benefit associated with a given type of technology. Often this takes the form of trying to understand why observed benefits are not as high as might be expected on the basis of the theoretical potential of a safety feature. As presented in more detail in several of the appendices of this report, factors that may account for such discrepancies can range from technical limitations in sensor technology to implementation details to human interface and interaction considerations. The background "Technology Review sheets" developed for each technology during the course of the project identify available information, hypotheses, research, and data that may bear on apparent effectiveness. Considering content across these Technology Review sheets effectively forms an underlying matrix of potential impact factors by technology.

Rating a Technology Class vs. Specific Vehicle Implementations

Early in the course of the project there was extensive discussion around whether the focus of the ratings should be on what would in essence be comparative safety benefit ratings of different technology types (e.g. lane departure warnings vs. adaptive headlights), or on specific implementations of a technology (e.g. lane departure warnings on model x vs. model y). While we consider the question of to what extent and how specific implementations of a given technology type vary in effectiveness to be of significant interest, a number of considerations led to the decision to focus for the present on a broad rating evaluation comparing different technology types. These considerations included:

- The directive in the RFP to rate across a range of technologies (possible examples suggested included back-up cameras, intelligent cruise control / adaptive cruise control, lane-departure warning, adaptive headlights). In-depth consideration both within a technology class and across multiple technology types was not practical within the time frame and resources of the current phase of the project. It was decided that once a base rating of relevant safety systems was established, this would provide a good foundation for future work examining whether and to what extent different implementations vary in effectiveness.
- Extensive model-by-model specific evaluations were clearly outside the scope of the project in terms of time frame and resources. Moreover, the National Highway Traffic Safety Administration's (NHTSA) expanding New Car Assessment Program (NCAP) provides the consumer with a resource listing for obtaining information on whether various vehicle models do or do not offer selected safety technologies that meet a minimum level of performance. The Insurance Institute for Highway Safety (IIHS) has recently expanded their testing programs to begin to consider model specific performance of safety technologies, beginning with scenario specific test track evaluations of forward collision braking / mitigation systems. It was concluded that developing this project in ways that complement rather than attempt to duplicate these undertakings was in order.
- From a conceptual standpoint, the most appropriate assessment of actual safety benefit would be based on epidemiological data examining the extent to which a given technology impacts crashes, injuries, and/or fatalities in the actual driving population. While the total number of adverse events per year is unacceptably high, the number of events per individual vehicle model, identifiable as being equipped with and without a given technology, is relatively low for purposes of calculating effect statistics. Thus, initial development of the rating system focusing more broadly on a class of technology rather than attempting to make evaluations across all models again seemed the most prudent starting point.

Considering the above, a reasonable question that might be asked is "How is this rating system going to help a consumer choose one car over another?" In brief, the answer is that the top level rating system is not intended as a car-buying guide, but rather is intended to serve as a technology-buying guide. It is intended to assist the consumer in identifying safety technologies that they may wish to look for in their next vehicle or consider as options within specific vehicle models of interest to them. After a consumer identifies a safety technology of interest to them, resources such as NCAP listings and IIHS evaluations provide a means to obtain vehicle / model specific information. Again, this does not preclude continuing / expanding the project over time to delve deeper into features or implementation characteristics that objective data suggest impact the overall effectiveness of a type of technology.

Selected Materials Developed During the Course of the Project

A range of materials have been developed during the course of the project that can be utilized in crafting suitable mechanisms for sharing information and findings with consumers. Key materials are identified below.

Factor Identification: Extensive consideration went into the identification of factors with potential relevance to evaluating the effectiveness of various safety technologies. Several interim documents developed during this process are reproduced in Appendix E. While a reduced grouping of factors was eventually employed in the development of the individual Technology Review sheets, some of the concepts and details considered in these interim documents may prove useful in future, further work on the evaluation of specific technologies. With this in mind, these selected materials are preserved in this report for reference purposes.

Research Notebooks: Our survey of the existing literature and data sources on selected safety technologies began with the development of research notebooks. These notebooks consist largely of "clippings" and notes collected on each technology along with a listing of sources of the material. Where particularly relevant, summaries of feedback received from industry representatives in response to technology rating forms and other requests were integrated into the notebooks using coded reference to industry sources to maintain appropriate anonymity. These notebooks merely provide raw background material; they are not polished as formal documents.

The layout of the notebooks largely follows the long version of the technology information form (Appendix G) that was sent to industry sources who expressed a willingness to contribute to the information gathering process.

Technology Review sheets: The technology notebooks served as the starting source for the development of summary Technology Review sheets on each technology. The Technology Review sheets organize material into the following categories:

- *What is the technology?* – A brief description of the technology.
- *Crash Reduction/Prevention* – A summary of key findings relating to both estimated and observed reductions in events relevant to evaluating the safety benefit of the technology. These are the values that are considered in the consumer oriented, top level rating of safety benefit.
- *Consumer Awareness & Trust* – A summary of identified information on the extent to which consumers are aware of the technology, understand how it is intended to function, and the degree to which they trust / have confidence in the technology. Much of the data in this section is based on self-report surveys and should be interpreted within that context.
- *Mobility Significance* – This section was originally intended to capture data that might be particularly relevant to the extent to which the technology might enhance older driver mobility. At present, it is being used to capture information that might be relevant to the extent to which a technology appears to enhance the mobility of any individuals who might otherwise be limited in their driving due to age or other sources of limitation.

- *Other Benefits* – A section for capturing other benefit-relevant information not otherwise covered.
- *Technology Penetration* – Considers the extent to which the technology is readily available and/or is present in the active vehicle fleet.
- *Frequency of Use* – Considers the extent to which drivers actually use a given technology, if they need to actively engage the technology, or if the option to turn the technology off is available.
- *Training and Education* – Some systems require little or no familiarity with the technology to derive benefit from them. Others have a steep learning curve for a user to become comfortable with them, but may become second nature once the user has developed a good mental model of how they work. This section considers to what extent the user needs to learn how to use the system to derive benefit and whether there is any identified data on how long most users take to become comfortable with the technology.
- *Behavior Adaptation* – This refers to behavior changes resulting from the use of the technology that may impact its net safety gain.
- *Auditory / Visual / Haptic / and Cognitive Demand* – A technology may offer potential benefits while also placing certain demands on the driver before they can derive that benefit. Engaging a system may involve a degree of mental, visual, manipulative, or auditory workload. Attending to a warning may similarly require some amount of attention and resource allocation that may impact effectiveness or even introduce distraction into the driving situation. This section considers what data are available on the extent to which the technology places some level of demand on the driver in each of the listed domains.
- *Vehicle Type* – This section considers the extent to which a particular technology may be more relevant to a particular type of vehicle. For example, electronic stability control (ESC) has been found to have a much more significant impact on overall safety gains in vehicles that are inherently less stable due to high centers of gravity (e.g. many SUVs) than in vehicles that are inherently more stable.
- *Limitations / Failure Conditions* – This section considers conditions under which the technology will not operate, performance may degrade, or actual failure may occur (i.e. weather, speed, tolerance boundaries, etc.)
- *Differences between Implementations* – This section considers whether there are known major differences between implementations of this general class of "technology" that need to be considered in evaluating its effectiveness, understanding or using the technology.
- *References* – Sources of information and data incorporated into the technology review sheet as well as other key references consulted but not directly cited.

As noted above, the *Crash Reduction/Prevention* section lists the sources from which objective values have been drawn to rate the safety benefit of the technology. Other sections cover material that is relevant for developing educational support material for the consumer. Many of the sections highlight factors that are highly relevant for both system designers and applied researchers to consider in identifying areas and features that might be improved to derive further gains in overall safety benefit in a technology class or specific implementation. It can be readily noted in working through the Technology Review sheets that the availability of objective data on many of these factor areas is relatively sparse or missing altogether. Areas where relevant objective data appears to be limited or missing

suggest areas of potential research needs. Due to the central nature of the Technology Review sheets to the project, copies of the current versions are included as appendices in this report.
The Technology Review sheets are seen as dynamic documents that should ideally be updated in an ongoing basis as additional research and associated data becomes available on the technology class and as implementations of the technology evolve. Current versions of the Technology Review sheets are reproduced in Appendix C.

<u>Educational Support Material:</u> The educational support material sheets include the following content sections:

- *What Is It?* – A short description of what a technology is and the conditions under which it might be relevant. This is intended as a very brief, high-level orientation to the technology.

The sections listed below represent a next level down description and elaboration of information on the technology, but again presented at a consumer-oriented level.

- *Why Would I Use This Technology?* – This section generally expands somewhat on what the technology is, why it is relevant, and sometimes includes additional information on how and/or why it works.
- *What Do Drivers Think?* – This section generally highlights information on consumer satisfaction with the technology.
- *How Well Does It Work?* – This section addresses objective data on the expected potential and/or observed safety benefits of the technology. It is based on selected data drawn from the *Crash Reduction/Prevention* section of the technology review sheet.
- *Who Benefits Most?* – This section highlights relevant information on the type of drivers, driving conditions, or type of vehicles that may benefit the most from availability of this technology.
- *In What Situations Doesn't It Work*? - Similar to the *Limitations / Failure Conditions* section of the technology review sheet, but presented at a consumer level.
- *Mobility Significance* – Similar to the *Mobility Significance* section of the technology review sheet, but presented at a consumer level.
- *Not All Systems Are Alike* – Similar to the *Differences between Implementations* section of the technology review sheet, but presented at a consumer level.
- *Different Names, Same Idea* – A listing of alternate names that different manufacturers may use to describe their implementation(s) of a particular type of safety technology.

Key Concepts in Proposed Rating System

The next several sections detail essential concepts utilized in the proposed rating system and specify the proposed scaling for determining actual rating values.

- The proposed rating system is based on a 5 level scale similar to the familiar 5-star rating systems employed in a variety of consumer familiar contexts.
  - As a working model, we are currently employing a geometric star symbol. This means that ratings will range on a scale from 1 (★) to 5 (★★★★★).

- The system is intended to provide an evaluation of a general class of technology, not to rate specific vehicle implementations.
  - The extent to which there is a significant difference in effectiveness across vehicle implementations or significance based on vehicle type [think of the benefit gain with ESC in SUVs vs. for vehicle frames that are inherently more stable to begin with] will be addressed in the deeper level educational / support information developed as part of the project.

- The rating for a given technology will reflect the high end of what "good" data indicates the probable benefit of the technology can currently provide (i.e. it will reflect the upper end of possible benefit).
  - Again, the intent is to assist the consumer in identifying potential benefits that may be associated with a technology to assist them in identifying technologies that they may wish to investigate further and encourage consideration in buying decisions that are relevant to their particular driving and life situation.

Projected vs. Demonstrated Benefit

It takes many years to develop objective data on the extent to which benefits are (or are not) observed with a given technology in the real world. In the interim, it is important to provide consumers with some guidance on "new" technologies, for which epidemiological data are not yet available. We believe it is important to promote "projected safety benefit" while being truthful to the consumer regarding demonstrated value.

To represent the distinction between projected vs. demonstrated benefit, two forms of the star symbol have been proposed:

- Open star (☆) ratings representing best-case estimates of projected or theoretical benefit based on simulation, test track, and/or experimental field data; and
- Solid star ratings (★) representing best-case estimates based on real-world demonstration of benefit drawing on field-operational tests, naturalistic data, and epidemiological and/or actuarial data.

There has been extensive discussion around whether the open vs. closed star symbolism is useful in making the distinction between projected/theoretical benefit and objectively demonstrated benefit in the driving population. A number of panel members have expressed concern that this distinction might be confusing for some consumers. The project leads continue to feel that this representation has conceptual value, but recognize that further work with communications experts and consumers would be required to resolve this question. In addition, as the rating of technologies in and out of the car continues to evolve, engagement with communications experts may also be useful in assessing if consumers might benefit from the eventual development of a more graphically complex, but information rich rating structure that could integrate further information about a technology's performance and limitations. For now, the open and closed symbol approach, with several alternate methods of representing this distinction in graphic presentation of the rating matrix concept, is presented later in this report.

Rating Types: Overall Safety Benefit & Scenario Specific

One of the challenges that we have spent significant time considering is how to best represent the safety benefit of one technology relative to another. In one sense, a technology that offers the potential to save the largest absolute number of lives should logically receive a higher rating than a technology that offers the potential to save a much smaller number of lives. On the other hand, what if the first technology is relevant to a large percentage of all possible crash events, but only actually works successfully in a modest percentage of those cases – while the second technology is designed to function in a much more limited number of situations, but is highly successful in presenting loss of life under those conditions. It thus seems "unfair" in a sense to down-rate the second technology relative to the first. This may particularly be the case if the scenario that the second technology is designed to mitigate or eliminate is of particular concern or relevance to a particular consumer or class of consumer.

This led to the proposal to rate technologies both in terms of **Overall Safety Benefit** (considering the maximum number of lives, injuries, or crash events) and in terms of benefit within **Specific Scenarios** the technology was designed to address. Thus, the top level rating matrix presented in this document includes ratings considering technologies from both perspectives.

Scaling Details

Ratings along the 5-star levels are based on a percentage of cases a particular technology has the potential to, or has been observed to, impact. The literature most frequently considers impact in terms of reduction in fatalities, frequency and/or severity of injuries, or in numbers of crash events.

We had at one time considered several approaches for calculating the relative value of a fatality vs. an injury vs. property damage associated with a crash. While there clearly is some value in such calculations for particular purposes, it soon became apparent that attempting to justify any such approach was likely to detract from the overall focus of the rating system and significantly added to both the complexity and actual validity of any calculations. We have therefore proposed treating each of the categories **Fatalities**, **Injuries**, and **Crashes** individually.

- Where percentage reduction values can be derived for any given category, it is considered in the evaluation.

- If percentage values are available in more than one category, the category showing the highest percentage benefit is used in the benefit level calculation. (This goes along with the recommendation by a number of Panel members that the best-case evaluation for each technology be considered in the rating.)

In addition to the three primary categories, insurance claim related data is available for some technologies that may contribute additional objective data on real-world impact. Many of the values currently available for emerging safety technologies are based on relatively limited numbers of cases or can be difficult to interpret for a number of technical reasons. When such data is available, such as through analyses released by the Highway Loss Data Institute (HLDI), such data may be considered in estimates of the technology's impact. It has been suggested that insurance claims may be a more complete tally of crashes than the more traditional reliance on police reports in that this data can include events that are recorded by the police as well as events that are not. Similarly, a HLDI representative commented that estimated effects based on injury claims in their published reports are most likely a reasonable representation of the injury events. While the majority of such events would be expected to be reported to the police, the insurance data may again pick-up some that are not or that are missed by police on the scene.

Scaling Theoretical vs. Observed Benefits

As noted previously, early in a safety technology's deployment, little or no real-world actuarial data will be available to realistically evaluate the actual extent to which a technology is producing a real safety benefit. Generally, actual benefit is found to be somewhat lower than potential benefit due to a range of factors such as unexpected technical limitations, driver override of a technology, and behavioral adaption, among others. Since actual observed benefits are expected to be lower than early estimated benefits, the rating system requires a higher theoretically estimated percentage benefit value (☆) to obtain a given star level rating than to obtain the same rating based on observed real-world benefits (★) (see Table 1).

Scaling Overall Benefit vs. Scenario Specific Benefits

Given the unacceptably high number of adverse events on the nation's roadways, even a relatively modest percentage reduction in total events is quite meaningful. Estimated annual crashes involving passenger vehicles per year derived from the period 2004-2008 averaged 5,825,000 crashes, 698,000 nonfatal injury crashes, and 33,035 fatal crashes (Jermakian, 2011). On the other hand, if a technology is designed to address a specific event scenario, a somewhat higher percentage impact would be expected to obtain an equivalent scenario specific benefit rating (see Table 1).

Proposed Scaling Values

Table 1. Conceptual Approach to Scaling for Overall Safety & Scenario Specific Benefits

	Overall Benefit		**Scenario Specific**	
	Theory, Sim., TT	FOT / Actuarial	Theory, Sim., TT.	FOT / Actuarial
Level 1	1 to 6%	1 to 5%	1 to 12%	1 to 10%
Level 2	7 to 14%	6 to 11%	13 to 28%	11 to 22%
Level 3	15 to 25%	12 to 19%	29 to 50%	23 to 38%
Level 4	26 to 40%	20 to 30%	51 to 80%	39 to 60%
Level 5	=> 41%	=> 31%	=> 81%	=> 61%

Sim. = Simulation, TT = Test Track evaluation, FOT = Field Operational Tests

The implications of the proposed scaling values presented above are best reviewed by considering how they translate into actual ratings of the technologies based on currently identified data. This is covered in the next section and in the proposed rating matrix that follows. (The rational structure for scaling is detailed in Appendix A.)

Extracted Values from Rating Sources

In each of the rating sections, we first consider the theoretical or best-case estimate of the potential benefit of a particular technology. A best-case estimate might be adjusted somewhat based upon "expert opinion" if multiple estimates are available and/or there are sound grounds for otherwise adjusting a value. In instances when reported values are adjusted, a rational for the adjustment is provided. Based upon the resulting values, an overall potential benefit rating is assigned both for a specific scenario for which the technology is designed to address, and a benefit rating relative to all vehicle related crash events.

Notes:

- All crash reduction estimates are based on estimated impact on anticipated rates in the United States.
- The sources cited in the following sections represent highlighted values used to determine a best-case rating for each technology. The full set of references considered in developing the ratings are provided in the individual *Technology Review sheets*.

Reference Technologies

Two technologies were considered as part of this rating system project largely to provide reference points for the development of the rating scale. These are Electronic Stability Control (ESC) and Adaptive Cruise Control (ACC) (sometimes referred to as intelligent cruise control and other related terms).

As detailed more fully in other portions of the deliverables for this project, ESC represents a fairly mature class of technology for which a reasonably robust body of real-world data exists on which to objectively evaluate the real-world safety benefits. These benefits have been deemed sufficiently substantive that NHTSA established Federal Motor Vehicle Safety Standard No. 126 to require ESC systems on all new production passenger vehicles, trucks, and buses with gross vehicle weight ratings of 10,000 lbs. or less by September 1, 2012. As documented in the ESC section, this technology has been established as providing a high safety benefit both for the specific scenarios for which it was designed, as well as impacting a substantial percentage of the overall risk associated with motor vehicle travel. Consequently, it was seen as a technology that should represent a top level rating both at the scenario specific level and a top level rating in terms of overall safety benefits.

At the other end of the continuum, ACC is seen as a technology that was developed largely as a convenience feature that has been considered as offering some modest potential safety benefit under limited circumstances. Thus, while ACC might be seen as offering relatively high value as a convenience feature, its relative rating as a safety technology might reasonably be expected to fall at the lower end of a safety benefit scale.

Electronic Stability Control (ESC)

Initial Ratings	**Overall**	**Scenario Specific**
Potential Overall Benefit	☆☆☆☆☆	☆☆☆☆☆
Benefit Currently Documented	★★★★★	★★★★★

*See individual **Technology Reviews** in **Appendix C** for full listing of data sources considered, including extracted values and full citations. While the term "ESC" is well established in the literature, a variety of manufacturer specific names are used for the technology class; these include: Vehicle Stability Assist, Vehicle Dynamic Control, Electronic Stability Program, Dynamic Stability Control, StabiliTrak, AdvanceTrac, etc.*

A substantive body of research is available to assess both the potential and demonstrated real-world benefit of ESC technology, and this is a primary reason for including ESC as a reference point in the proposed rating system. See the ESC *Technology Review* in *Appendix C* for citation details and an extended overview of available data. In one representative study, an IIHS report (Farmer, 2010) based on 10 years of data in the Fatality Analysis Reporting System (FARS) found an overall reduced fatal crash involvement risk of 33 percent for ESC-equipped vehicles. This would translate into a 5 solid star rating in the Overall Benefit category using the proposed scaling values presented earlier in Table 1. In cases where the objectively demonstrated benefit of a technology is already at the highest ranking on the scale, this in effect provides the most meaningful evaluation of "potential," and the 5 open star ranking is assigned as well.

As presented in the ESC *Technology Review*, ratings for various Scenario Specific benefits of ESC technology range from 56 to 90 percent in the fatalities reduction category, depending upon the vehicle type considered. Objective data is available for injury reduction at 70 percent, and crash rate reductions at the 72 percent level have been reported. These values place the Scenario Specific rating for ESC at the 5 solid star rating level.

Adaptive Cruise Control (ACC)

Initial Ratings	**Overall**	**Scenario Specific**
Potential Overall Benefit	☆	☆☆
Benefit Currently Documented	★	★★

*See individual **Technology Reviews** in **Appendix C** for full listing of data sources considered, including extracted values and full citations. In addition to Adaptive Cruise Control, other brand-oriented terms, such as Autonomous Cruise Control and Intelligent Cruise Control, may be used.*

The one publically available report that was identified that provided an estimate of potential safety benefits of ACC was a NHTSA sponsored field operational test (FOT) (Koziol et al., 1999). The authors concluded that if such systems were fully deployed and utilized at the engagement rate seen in the FOT, it was estimated that the number of collisions on freeways for travel velocities above 40 km/h would be reduced by 17 percent for two specified scenarios (highway driving when vehicles are traveling over 40 km/h when an ACC-equipped vehicle approached a slower vehicle traveling at a constant velocity, and when a lead vehicle decelerated in front of an ACC-equipped vehicle). This estimate would correspond to a reduction in the number of police-reported rear-end collisions by about 13,000 in 1996, and this was interpreted as indicating a fairly strong benefit compared to manual driving. However, as a percentage of total crashes of all types, this would correspond to less than 1 percent. These FOT based estimates appear to provide the only substantive values that are available to work from for current rating purposes. While providing a limited basis for estimating benefit, the values are based on FOT based evaluation and qualify for solid star rating under the proposed system. This translates into qualifying for 2 solid stars under the Scenario Specific rating and 1 solid star under the Overall Benefit category.

It was noted in the report that additional safety benefits might be expected from a reduction in other rear-end collisions involving cut-ins and lane changes and from use of ACC on roadways other than freeways; however, benefit estimates for these scenarios were not examined in the FOT. Drivers were found to engage the system for 6 percent of the time on arterials and 11 percent on state highways, thus limiting the percentage of time during which potential benefits might be realized.

More recent work, such as studies carried out by IIHS, largely consider vehicles that frequently combine ACC with forward collision warning (FCW) and, increasingly, autobrake features. Personal communication with IIHS personnel indicates that they see it as difficult to isolate the effects of ACC from the other components of these systems in work going forward.

Primary Technologies
Adaptive Headlights

Initial Ratings	**Overall**	**Scenario Specific**
Potential Overall Benefit	☆☆	☆☆☆☆☆
Benefit Currently Documented	★	★★★

See individual ***Technology Reviews*** *in* ***Appendix C*** *for full listing of data sources considered, including extracted values and full citations. Note that the term 'adaptive headlights' can be used to refer to a range of technologies. The term is used here to apply largely to headlights that adjust the angle of illumination taking into consideration steering, speed, and elevation of the car. In addition to "adaptive headlights," these might also be referred to as "curve illuminating," "steerable headlights," etc.*

The potential benefit ratings for adaptive headlights are based largely on an IIHS analysis by Jermakian (2011). Considering scenarios related to improving visibility when negotiating curves in darkness or twilight, the study estimated that adaptive headlights have theoretical relevance to 90 percent of the crashes that occur on curves at night, 91 percent for nonfatal injury crashes, and 88 percent for fatal crashes. This would translate into an overall maximum safety benefit potential across all event types of: 8 percent for fatalities, 4 percent for injuries, and 2 percent for crashes. While the estimates in the Jermakian study do include some adjustments for known limitations of then-current systems, they still do largely represent theoretical maximum benefits. It seems most appropriate at this time to use these values in an open star configuration to represent system potential. Using the Scenario Specific injury benefit value of 91 percent reduction, a 5 open star rating can be applied. At the Overall Safety Benefit level, the 8 percent reduction in fatalities value translates into a 2 open star rating.

A HLDI analysis looked at adaptive headlights offered by Acura, Mazda, Mercedes, and Volvo and found that property damage liability claims fell in the 5 to 10 percent range for vehicles with adaptive headlights compared to vehicles without adaptive headlights (IIHS, 2012). This strongly suggests a real-world benefit being realized in vehicles equipped with this technology, although it is difficult to extrapolate this value into an objective percentage value for the fatalities, injuries, and crash categories. *For purposes of the Overall Benefit category, we are proposing treating this high-end 10 percent reduction in insurance claims value as a surrogate for one of the three standard categories (fatalities, injuries, and crashes). Discussions with IIHS (Lund, 2014) indicate that this number would best be translated into a high-end estimate of around a 2.5 to 5 percent reduction in overall crash events; i.e., there are on the order of two property damage liability claims for a crash event that involves two vehicles.* Using this value, this results in assigning a 1 solid star rating in the Overall Benefit category. *(It should be noted that the HLDI reports include estimates for reductions on collision claims that represent crashes resulting in damage to the insured vehicle that is not the fault of the driver of another vehicle.)*

It is an open question as to how best to translate the aforementioned values into a scenario specific rating since the insurance data is not necessarily limited to the events occurring on curved roadways and night driving. On the other hand, the apparent best-case reduction estimates in the currently available insurance data are in the same general range of the total number of scenario specific events that might be anticipated. Therefore, for purposes of the current proposal, we have conservatively assigned a Scenario Specific rating of 3 solid stars. It may be appropriate to reconsider in the future whether the scenario rating for adaptive headlights should be expanded to consider twilight and nighttime driving in general as opposed to the more restrictive scenario of driving on curved roads in low-light and dark driving conditions.

Back-Up / Rear-View Cameras

Initial Ratings	**Overall**	**Scenario Specific**
Potential Overall Benefit	☆	☆☆☆
Benefit Currently Documented	★	★★

*See individual **Technology Reviews** in **Appendix C** for full listing of data sources considered, including extracted values and full citations.*

While providing a back-up camera to increase a driver's ability to see directly behind the vehicle is an intuitively appealing concept, relatively limited data is available on the extent to which such systems actually provide a net benefit. Across a series of experimental studies, Mazzae (2008, 2010, 2013) reported that the use of back-up cameras reduced crashes in an unexpected collision trial by approximately 30 percent. Treating these experimental studies as providing a theoretical estimate of potential benefit, this would translate into a scenario specific rating of 3 open symbols.

Real-world performance data is similarly limited at this time. Two studies from the HLDI (2011, 2012) consider initial data on the impact of the presence of back-up cameras on insurance claims and damages. The data for Mercedes-Benz vehicles showed small and mixed findings across insurance claims and damages. The report concluded that the data showed no significant effect on any insurance coverage; however, this was considered a relatively weak analysis for injury effects involving pedestrians, and it was stated that additional analyses were underway. The data for Mazda vehicles found that, contrary to expectations, there was an increase in collision frequency claims (3.1%), severity, and overall losses ($18), but a non-significant reduction in property damage / liability claims. Most relevant from a safety perspective, there was a reduction in the frequency of high severity bodily injury claims of 22.2 percent, although the overall frequency of bodily injury claims was down a non-statistically significant 3.1 percent.

If the 22 percent value for reduction in high severity bodily injury for the aforementioned vehicle type in is used as a measure of real-world potential, then a rating of 2 solid stars could be applied under the Scenario Specific category, resulting in a combined rating of 2 solid and 3 open stars.

Note: It should again be emphasized that the relatively modest ranking for back-up cameras in terms of Overall Safety Benefit is a result of the relatively low number of backup event related injuries and fatalities relative to the total number of driving related injuries and fatalities. It is also recognized that there is a particularly high emotional cost associated with this type of event. NHTSA (2014) estimates that 31 percent of all backup event fatalities involve children under five years of age, and another 26 percent are adults 70 years and *older; these events often involve family members or other close associations. Societal pressure to do something about such events led to NHTSA issuing a final rule on March 31, 2014 mandating rear visibility technology in all new vehicles under 10,000 pounds by May 2018.*

Forward Collision Warning (FCW)

Initial Ratings	**Overall**	**Scenario Specific**
Potential Overall Benefit	☆☆☆	☆☆☆☆☆
Benefit Currently Documented	★	★★

*See individual **Technology Reviews** in **Appendix C** for full listing of data sources considered, including extracted values and full citations. (Forward Collision Mitigation [FCM] systems that actively brake the vehicle are considered separately; see next section.) The terms Crash Imminent Warning and Pre-Crash Warning are sometimes used to describe this technology class.*

The IIHS analysis by Jermakian (2011) again provides a comprehensive assessment of potential safety benefit for this technology. However, a significant drawback of this study for our purposes is that it considers together both forward collision warning (FCW) and forward collision mitigation (FCM) / autobraking systems. As a consequence, the theoretical potential ratings for FCW alone may be somewhat elevated based on this data. Keeping that in mind, Overall Benefit estimates for potential fatality reduction of 17 percent (5,633 cases out of 33,035), non-fatal injury reduction of 21 percent (146,000 cases out of 698,000), and crash rate reduction of 25 percent (1,453,000 cases out of 5,825,000) are given. (Case counts are based on combining annual relevant front-rear crashes and relevant single vehicle crashes.) The latter would translate into an open star theoretical potential rating at the 3 open star level.

In a field operational test, 66 drivers were evaluated for four weeks each. Based on the number of near-crash scenarios identified, the system was projected to reduce rear-end collision rates by 10 percent (Najm, Stearns, Howarth, Koopmann, & Hitz, 2006). If this limited data set is used as a best-case estimate of existing real-world data, then this 10 percent reduction estimate would qualify for 1 solid star in the Scenario Specific rating category. IIHS has reported (Lund, 2013; 2014) that insurance data on property damage loss claims show a reduction of 5 to 7 percent for vehicles with FCW and suggest that this translates into a 10 to 15 percent reduction in rear crashes. Using this insurance data based estimate would increase the Scenario Specific rating category to 2 solid stars. Combining these values would produce an Overall Benefit rating of 1 solid star and 3 open stars, and a Scenario Specific rating of 2 solid stars and 4 open stars.

Note: As this report was being finalized, HLDI (2014) released a new bulletin evaluating insurance data on a FCW system paired with lane departure warning (LDW) technology in the Honda Accord. These data may provide grounds for upgrading the demonstrated benefit rating of one or both of these technologies.

Forward Collision Mitigation (FCM) / Collision Imminent Braking (FCB) / Autobrake

Initial Ratings	**Overall**	**Scenario Specific**

Potential Overall Benefit	☆☆☆	☆☆☆☆☆
Benefit Currently Documented	★★	★★★

*See individual **Technology Reviews** in **Appendix C** for full listing of data sources considered, including extracted values and full citations. In addition to the terms above, this technology class is sometimes identified by the following terms: Crash Imminent Braking, Autonomous Emergency Braking, Emergency Brake Assist, and Predictive Brake Assist.*

As discussed in the previous section, the current theoretical overall potential benefit rating for forward collision mitigation (FCM) / collision imminent braking / autobraking technology is based largely on the same estimates developed for FCW (Jermakian, 2011), which translates into a 3 open star rating. However, we are proposing that the theoretical values for the scenario specific rating be increased from 4 open stars (51 to 80% reduction) to a 5 open star rating (>80%) due to recently released IIHS test track data (IIHS, 2013) demonstrating that some vehicle types under specified scenarios are able to avoid or mitigate front-to-rear crashes greater than 80 percent of the time.

As noted in the *Technology Review* in *Appendix C*, implementations of forward collision mitigation technologies may include forward collision warning (FCW) and brake assist technology that pre-primes the brake system, in addition to actual autonomous braking capability. The present rating includes consideration of hybrid FCM systems that bundle FCW and/or brake assist technologies. This decision is largely pragmatic since most of the available real-world effectiveness data is based on such bundled technologies. It should be recognized that implementations that do not include warning and/or brake assist technologies may not provide the same level of benefit.

An insurance claims based study based on comparable Volvo models with and without a FCM system (Isaksson-Hellman & Lindman, 2012) reported a 23 percent reduction in crashes for the equipped vehicles. Based on current scaling, this would qualify as a best-case, scenario specific objective rating of 3 solid stars. As a percentage of total crash events, this would translate into a reduction of approximately 6.6 percent based on the proportions reported in Jermakian (2011); this would qualify as a best-case estimate of 2 solid stars. IIHS (Lund, 2014) reports similar reductions for vehicles equipped with FCW and autobrake in property damage loss data appearing as a 10 to 15 percent reduction in claims, which is interpreted as translating into a 20 to 30 percent reduction in rear crashes.

Lane Departure Warning (LDW)

Initial Ratings	**Overall**	**Scenario Specific**
Potential Overall Benefit	☆☆☆	☆☆☆
Benefit Currently Documented	★	★

*See individual **Technology Reviews** in **Appendix C** for full listing of data sources considered, including extracted values and full citations. (Note: this review does not consider "lane keeping assist" / lane keeping mitigation technologies that adjust steering to actively assist in keeping the vehicle within lane boundaries.)*

An IIHS analysis by Jermakian (2011) provided independent estimates for the maximum potential reductions across lane departures resulting in four types of scenarios: single vehicle crashes, head-on crashes, sideswipes between vehicles moving in the same direction, and sideswipes between vehicles moving in the opposite direction. Combining these lane departure events into a single generic scenario of crashes associated with lane departures, annual reduction totals of 179,000 crashes (out of 5,825,000), 38,000 non-fatal injury crashes (out of 698,000), and 7,529 fatal crashes (out of 33,035) translates into corresponding theoretical reductions of 3 percent, 5.4 percent, and 23 percent, respectively, in the Overall Benefit category. Using the 23 percent reduction value for fatal crashes, this translates into a 3 open star rating. At the scenario specific level, the highest substantive theoretical estimate that we have located is Jermakian's (2011) estimate indicating a possible 46 percent reduction in fatalities in lane departure associated head-on crashes. This also corresponds to 3 open stars in the Scenario Specific category.

Obtaining objective data on the extent to which LDW systems are providing real-world safety benefits has proven to be challenging. Extracting data from a field operational test of one specific system (Wilson, Stearns, Koopmann, & Yang, 2007) suggests possible reductions in scenario specific crashes in the range of 1-8 percent. The scenarios considered were more restrictive than the broader scenario categories considered by Jermakian (2011) and the high end 8 percent reduction might be considered an optimistic rating. If this value is applied, a 1 solid star rating could be assigned to the scenario category. This suggests a translation into an Overall Benefit level of at most 1 solid star at this time. Available actuarial data also suggest a modest, at best, assessment of LDW systems as a class at the current time. A recent assessment by IIHS (Lund, 2013) concluded that "Lane departure warning may have the potential to reduce fatal crashes, but so far no benefits from this feature have shown up in insurance data." The apparent discrepancies between theoretical estimates of LDW potential and presently available objective real-world data represents a case in point for the concept of a combined solid and open star rating system. Given the data identified to date, LDW would receive a rating of 1 solid and 3 open stars.

Note: As this report was being finalized, HLDI (2014) released a new bulletin evaluating insurance data on a forward collision warning (FCW) system paired with lane departure warning (LDW) technology in the Honda Accord. These data may provide grounds for upgrading the demonstrated benefit rating of one or both of these technologies.

Conceptual Presentation of the Rating Matrix

As detailed in the previous sections, we have proposed an approach to rating safety technologies that is based on a number of key elements. These include:

- The use of a 1-5 scale that consumers are familiar with from everyday rating systems;
- Distinguishing between projected / theoretical potential and demonstrated benefit of production level technologies under actual real-world driving conditions; and
- Consideration of the overall safety benefit of a given technology relative to all expected driving related crash, injury, or fatality events versus the safety benefit of the technology relative to the specific type of driving situation the technology has been designed to mitigate.

The latter two aspects of the proposed approach are seen as particularly important for supporting different types of "customers" in approaching the ratings in the way that is important to them. Individuals who are drawn to early adoption of technologies based largely on potential or who are willing to pay a premium to maximize possible safety gains prior to the full development of actuarial data, will find the ratings of technology potential the most relevant. On the other hand, individuals who are more comfortable adopting technologies only after they are well established and have a clearly demonstrated level of gain, will find the ratings based on demonstrated benefit most meaningful. Along another dimension, some individuals simply want everything reduced to a single consideration, i.e. what is going to have the "biggest" overall impact on risk? These types of consumers are likely going to be drawn to the "overall" benefit rating of the system. Conversely, a consumer who is willing to take a deeper look at their own driving situation in deciding what technologies are most relevant to them may well find the "scenario specific" rating aspect of the proposed system most useful. There is a clear challenge in developing a presentation approach that supports the interests of each of these types of consumers in a concise and relatively easy to understand manner.

During the course of this project, a number of upper-level approaches to summarizing the results of the proposed rating system have been explored. Feedback from the panel and industry representatives has generally pushed for simplifying the presentation of the ratings. The authors agree that simple, clear communication of ratings is the ideal goal; further work may wish to investigate the consumer's ability to "digest" rating structures that comprehensively represent various perspectives that consumers may wish to consider when evaluating the technologies to determine personal priorities. The following pages present four variant methods of presenting the upper-level of the assessment matrix at a conceptual level. (These are referred to as "conceptual" in that they are primarily intended to represent the content to be presented to the consumer – not necessarily the exact graphical layout that might be used. It is assumed that consumer oriented graphic design experts may well significantly improve upon the conceptual layouts presented here. Furthermore, it has always been assumed that web enabled forms of the matrix might well provide point and click links to deeper levels of detail, etc.) Version A represents an approach currently favored by the project leads. It uses the open and closed star ratings to highlight the difference between the potential benefit and observed benefit ratings. Versions B and C merge the open and closed star ratings into an overlapped presentation so that open stars are only visible if theoretical benefit is rated as exceeding currently demonstrated benefit. Version D differentiates theoretical from demonstrated benefit but drops the use of open stars to highlight these different dimensions.

A. Top Level Matrix with Initial Ratings – Potential & Observed Benefit Scaled Separately

Overall Safety Benefit (across all event types)	**Electronic Stability Control**	**Backup Cameras**	**Adaptive Headlights**	**Lane Departure Warning**	**Adaptive Cruise Control**	**Forward Collision Warning**	**Forward Collision Braking**
Potential Benefit	☆☆☆☆☆	☆[a]	☆☆	☆☆☆	☆	☆☆☆	☆☆☆
Benefit Currently Documented	★★★★★	★	★	★	★	★	★★

Scenario Specific Benefit	Loss of Steering Control	Back-Up Event	Dark Curves	Lane Departure	Rear End Collision		
Potential Benefit	☆☆☆☆☆	☆☆☆	☆☆☆☆☆	☆☆☆	☆☆	☆☆☆☆	☆☆☆☆☆
Benefit Currently Documented	★★★★★	★★	★★★	★	★★	★★	★★★

Ratings are based on currently identified data and percentage cut-points as defined under the draft working model; these ratings may be adjusted prior to any public release as data evaluation and system structure is refined based upon on-going feedback from the Expert Panel and other contributing sources.

All ratings are on a 1 to 5 scale. Open star (☆) ratings represent best case estimates of projected or theoretical benefit based on simulation, test track, and/or experimental field data. Solid star ratings (★) represent best case estimates based on real-world demonstration of benefit drawing on field-operational tests, naturalistic data, epidemiological and/or actuarial data.

Loss of Steering Control - Skidding on slippery surfaces, loss of traction with unexpected or high speed turn

Back-Up Event - Pedestrian fatality, injury or non-injury crash when backing–up a vehicle

Dark Curves - Impact of improved visibility when attempting to negotiate curves in darkness or twilight

Lane Departure - Unintended drift out of lane or failure to use turn signal to warn other drivers

Rear End Collision - Front of your vehicle with rear of a lead vehicle

[a]The relatively modest ranking for Back-Up cameras in terms of Overall Safety Benefit is a result of the relatively low number of backup event related injuries and fatalities relative to the total number of driving related injuries and fatalities.

B. Top Level Matrix with Initial Ratings – Combined Scaling with Scenario Detail

Technology Rating Type	Electronic Stability Control	Backup Cameras	Adaptive Headlights	Lane Departure Warning	Adaptive Cruise Control	Forward Collision Warning	Forward Collision Braking
Overall Safety Benefit (across all event types)	★★★★★	★[1]	★	★☆☆	★	★☆☆	★★☆
SPECIFIC SENARIOS							
Loss of Steering Control Skidding on slippery surfaces, loss of traction with unexpected or high speed turn	★★★★★	-	-	-	-	-	-
Rear End Collision Front of your vehicle with rear of a lead vehicle	-	-	-	-	★★	★★☆☆	★★★☆☆
Lane Departure Event Unintended drift out of lane / failure to use turn signal to warn others resulting in an event	-	-	-	★☆☆	-		
Dark Curves – Impact of improved visibility when attempting to negotiate curves in darkness or twilight	-	-	★★★☆☆	-	-	-	-
Back-up Event Pedestrian fatality, injury or non-injury crash when backing –up a vehicle.	-	★★☆	-	-	-	-	-

Ratings above are based on currently identified data and percentage cut-points as defined under the draft working model; these ratings may be adjusted prior to any public release as data evaluation and system structure is refined based upon on-going feedback from the Expert Panel and other contributing sources.

All ratings are on a 1 to 5 scale. Open star (☆) ratings represent best case estimates of projected or theoretical benefit based on simulation, test track, and/or experimental field data. Solid star ratings (★) represent best case estimates based on real-world demonstration of benefit drawing on field-operational tests, naturalistic data, epidemiological and/or actuarial data.

[1]The relatively modest ranking for Back-Up cameras in terms of Overall Safety Benefit is a result of the relatively low number of backup event related injuries and fatalities relative to the total number of driving related injuries and fatalities.

C. Top Level Matrix with Initial Ratings – Combined Scaling of Potential & Observed Benefit

	Technology						
	Electronic Stability Control	**Backup Cameras**	**Adaptive Headlights**	**Lane Departure Warning**	**Adaptive Cruise Control**	**Forward Collision Warning**	**Forward Collision Braking**
Overall Safety Benefit[1] (across all event types)	★★★★★	★[a]	★	★☆☆	★	★☆☆	★★☆
	Loss of Steering Control	Back-Up Event	Dark Curves	Lane Departure	Rear End Collision		
Scenario Specific Benefit[1]	★★★★★	★★☆	★★★☆☆	★☆☆	★★	★★☆☆	★★★☆☆

Ratings are based on currently identified data and percentage cut-points as defined under the draft working model; these ratings may be adjusted prior to any public release as data evaluation and system structure is refined based upon on-going feedback from the Expert Panel and other contributing sources.

All ratings are on a 1 to 5 scale. Open star (☆) ratings represent best case estimates of projected or theoretical benefit based on simulation, test track, and/or experimental field data. Solid star ratings (★) represent best case estimates based on real-world demonstration of benefit drawing on field-operational tests, naturalistic data, epidemiological and/or actuarial data.

Loss of Steering Control - Skidding on slippery surfaces, loss of traction with unexpected or high speed turn

Back-Up Event - Pedestrian fatality, injury or non-injury crash when backing–up a vehicle

Dark Curves - Impact of improved visibility when attempting to negotiate curves in darkness or twilight

Lane Departure - Unintended drift out of lane or failure to use turn signal to warn other drivers

Rear End Collision - Front of your vehicle with rear of a lead vehicle

[1]Benefit ratings reflect best case evaluations of existing systems. Not all implementations may offer the same level of benefit. Consumers may wish to consult the U.S. Government NCAP ratings for vehicle models offering select technologies meeting minimum performance standards, Insurance Institute of Highways Safety (IIHS) ratings of individual vehicle models, or other vehicle specific ratings for technologies of interest.

[a]The relatively modest ranking for Back-Up cameras in terms of Overall Safety Benefit is a result of the relatively low number of backup event related injuries and fatalities relative to the total number of driving related injuries and fatalities.

D. Top Level Matrix with Initial Ratings – Potential & Observed Benefit Scaled Separately (All Solid Stars)

Overall Safety Benefit (across all event types)	**Electronic Stability Control**	**Backup Cameras**	**Adaptive Headlights**	**Lane Departure Warning**	**Adaptive Cruise Control**	**Forward Collision Warning**	**Forward Collision Braking**
Potential Benefit	★★★★★	★[a]	★★	★★★	★	★★★	★★★
Benefit Currently Documented	★★★★★	★	★	★	★	★	★★

Scenario Specific Benefit	Loss of Steering Control	Back-Up Event	Dark Curves	Lane Departure	Rear End Collision		
Potential Benefit	★★★★★	★★★	★★★★★	★★★	★★	★★★★	★★★★★
Benefit Currently Documented	★★★★★	★★	★★★	★	★★	★★	★★★

Ratings are based on currently identified data and percentage cut-points as defined under the draft working model; these ratings may be adjusted prior to any public release as data evaluation and system structure is refined based upon on-going feedback from the Expert Panel and other contributing sources.

All ratings are on a 1 to 5 scale.

Loss of Steering Control - Skidding on slippery surfaces, loss of traction with unexpected or high speed turn

Back-Up Event - Pedestrian fatality, injury or non-injury crash when backing–up a vehicle

Dark Curves - Impact of improved visibility when attempting to negotiate curves in darkness or twilight

Lane Departure - Unintended drift out of lane or failure to use turn signal to warn other drivers

Rear End Collision - Front of your vehicle with rear of a lead vehicle

[a]The relatively modest ranking for Back-Up cameras in terms of Overall Safety Benefit is a result of the relatively low number of backup event related injuries and fatalities relative to the total number of driving related injuries and fatalities.

Initial Observations

Availability of Objective Data

It is quite clear that dramatic advances have been made in passive safety as evidenced in both crash test results and in observed increases in driver and passenger survivability when vehicles crash in the real world. As gains continue to be made in the passive safety domain, the industry is increasingly making investments in developing and marketing technologies that are intended to warn us of potential conflict situations, support our situational awareness of conditions around our vehicles, and even actively take limited control of the vehicle in apparent emergency situations. These are laudable investments and are ones that appear to offer great potential.

At the same time, perhaps the most concrete finding of this project to date is the observation of how relatively little objective data is available upon which to evaluate the real-world effectiveness of many types of this new class of safety technologies that are appearing at car dealerships and on our highways. Of the technologies considered in this report, only electronic stability control (ESC) can be classified as a technology for which we have sufficient real-world performance data from which to make a solid evaluation of its effectiveness, and that data clearly indicates a solid and substantial safety benefit. However, that is not to say that the underlying sensing and actuating technology of the other, newer systems have not undergone extensive research and development testing; there is reason to believe that extensive development and functional test level work goes into determining the basic technical capacities and limitations of individual components and basic system level function. The impression that comes from talking in-depth with individuals within the automotive industry is largely one of significant dedication to finding ways to make vehicles safer. At the same time, until a new technology is available and in use by the general public, it is striking to realize the extent to which relatively little can be established about how these systems actually perform outside of the laboratory or test track.

In some ways, this is not surprising. In contrast with advances in structural materials and engineered crumple zones, the effectiveness of many of these new safety technologies is likely to be dependent in part on how drivers interact, or fail to interact, with them. Comprehensive real-world testing / assessment of driver behavior is, in many ways, more complex and challenging. During the extensive interviews conducted as part of this project, a number of individuals from OEMs and tier one technology manufacturers made quite candid comments about where they look for information on how these systems actually perform in the vehicle fleet. Beyond subjective customer surveys that they conduct or commission, the industry largely looks to governmental agencies and organizations such IIHS / HLDI for the collection and analysis of crash and other event relevant data. We have found that the available data and analyses relevant to these newer technologies are even sparser than we anticipated going into this project. Other objective data from relatively real-world assessments such as field operational testing studies employing production level systems are also quite limited, as are relevant naturalistic studies. Consequently, the ratings presented in this initial assessment are based on a quite limited set of observed benefit data for all of the technologies relative to what is known about the performance of ESC.

In addition to our being stuck by the limited availability of objective real-world performance data, many representatives from industry, academics, and individuals from NGOs and other organizations that we spoke with were similarly surprised that more data could not be identified to provide objective demonstration of real-world performance. In reviews of various versions of the Technical Review sheets, a number of the aforementioned individuals frequently commented that they were sure there must be a study on "x" or that organization "y" had relevant data. Yet when asked for citations, few were typically found. Similarly, inquires to the organizations in question generally failed to turn up additional data. On the other hand, it should be noted that the experts that we questioned that came from epidemiological and database analysis backgrounds, particularly those that had in the past worked with governmental agencies charged with compiling such information, tended to be less surprised by the difficulty locating good data sources to use in the ratings.

While the finding that available data is currently quite limited is problematic in one sense, the way in which this was determined could in itself represent a constructive product of this phase of the project. As outlined in the introduction, and documented more fully in several of the appendices to this report, the search for available data involved a wide cross-section of professionals from the safety research, manufacturer and tier one supplier, academic, NGO, and governmental communities. Early discussions with identified experts, presentations before the Alliance of Automotive Manufacturers and the Association of Global Automakers, multiple broadly attended and industry wide teleconferences, numerous follow-up conversations, and the Advisory Panel meetings, all engaged a significant number of key domain professionals in far ranging consideration of these issues. This process thus stimulated discussion and encouraged thought within the safety research and manufacturer communities relative to developing a more comprehensive approach to thinking about safety and methods for objectively evaluating new technologies.

Current Scaling

As discussed earlier, the present scaling structure focuses on the concept of safety benefit in terms of reductions in crashes, injuries, and/or fatalities. ESC was selected as a reference technology that has sufficient penetration and history in the vehicle fleet to be used as an example of a highly beneficial technology that should reasonably score at the top rating level of the scale. The threshold for a top-level rating was set somewhat lower than the observed impact values for ESC. The remaining levels were scaled downward from that point using a rational scaling structure. The scaling is further structured so that higher impact values are required for theoretical estimates of safety potential than for actual observed impact in real-world data. Similarly, higher impact values are required when a technology is considered in terms of the specific scenario it was designed to work under as opposed to its impact relative to the total universe of crashes, injuries, and fatalities.

Using the initial set of scaling values, ESC ranks at the 5 star level in terms of Overall Safety Benefit (both theoretical and observed) and at the 5 star level in Scenario Specific Benefit (both theoretical and observed). Most of the other technologies considered rank relatively highly (3 to 5 stars) in terms of potential benefit within the scenario they were designed to impact (i.e. adaptive headlights, lane departure warning, forward collision warning, and forward collision braking). Only adaptive cruise control (ACC) ranks somewhat modestly in terms of potential benefit within its relevant scenario (rear end collision) and this is a function of ACC only being operational for a limited percentage of the time when it might potentially have a safety benefit (i.e. it is an option that a driver must actively engage). This result is reasonable when it is considered that ACC was designed primarily as a comfort or convenience system; safety advantages are really secondary benefits.

The scenario specific safety potential ratings for back-up cameras and lane departure warning systems do not obtain higher scores because of observed limitations of how drivers interact with these systems (see *Technology Reviews* for specific citations). Both experimental studies and surveys indicate that drivers frequently fail to actually use their back-up cameras and many often fail to notice objects in the back-up path even when orienting to the camera. These factors thus limit the theoretical gain expected from simply purchasing such a system. Lane departure warning systems are limited by the availability of visible lane markings and other technical considerations, and there is evidence that many drivers ignore or turn-off lane departure warning systems, again lowering the gain that might theoretically be expected of the technology. In addition, simulation data suggest that implementation differences – finding the right balance between warning drivers too early of a potential lane excursion (and risking driver frustration) and warning too late to have a significant reduction in actual risk – may play an important role in the real-world effectiveness of such systems. Similarly, the way in which warnings are delivered – auditory, haptic, etc. – may well merit significant investigation to assist in better understanding what appears to be something of a discrepancy between theoretical potential and observed actual benefits in clearly limited available real-world data. Future enhancements of these technologies may result in grounds for reevaluating this level of theoretical expectation. From a scenario specific potential benefit expectation, the scaling across the seven rated technologies appears fairly reasonable.

In terms of overall observed safety benefit, ESC is rated at 5 stars. Of the remaining six technologies, one (forward collision mitigation / autobraking) presently is ranked at 2 stars and the others all at 1 star. In terms of scenario specific benefits, of the remaining six technologies, two are rated at the 3 star level (Adaptive Headlights and Forward Collision Mitigation / Braking), three at the 2 star level (back-up cameras, adaptive cruise control, and forward collision warning) and one at the 1 star level (lane departure warning). It must be kept in mind that differences between potential benefit and observed benefit may in some or all cases be due to limitations in current system implementations. Careful review of the factors that might impact system effectiveness is clearly called for when differences appear between expected and observed ratings. Potential sources of explanation for some of the differences are identified in the relevant *Technology Reviews* in *Appendix C*. Targeted research on why apparent benefit is so modest seems particularly appropriate in technologies such as lane departure warning systems and back-up cameras. Adaptive headlights, forward collision warning and forward collision braking values may be relatively moderate at this time due in part to limited data availability; however, this strengthens the argument for why such real-world data collection is needed as adoption of some of these systems is being actively encouraged both by the automotive industry and government entities.

Given the issues just covered, the current observed benefit rating levels are not unreasonable, but it also seems appropriate to consider the present scaling as quite open for review. Responses from the Advisory Panel and industry representatives suggest that the scaling for observed data seems reasonable based on the logical structure of the scale and the relative deficit of objective data. We will continue to review the proposed scaling cut-points now that values from the technology review have been applied to the scale and as we continue to receive comments and suggestions from Advisory Panel members and industry representatives.

Limitations / Points to Keep in Mind

While significant effort has been made to present a set of ratings that realistically represents the current state of knowledge concerning the safety benefit of the technologies considered, we, in many ways, see this as a proposal for a rating system. That is to say, it organizes available information in a way that allows for serious review and discussion around what is currently known and not known about these emerging safety technologies. By presenting a rationale for how technologies might be rated, summarizing the data that was identified as being available to base ratings upon, and proposing an initial set of scaling values to use for making ratings, interested parties have something concrete to which to react. Since the charge for this project was to focus on data rather than opinion, individuals or entities that feel particular technologies are either under or overrated relative to others are encouraged to identify data that can be used to justify adjusting the current proposed ratings. As noted elsewhere, we particularly see the apparent discrepancies between theoretically projected benefit and observed benefit as a means to potentially focus attention on and motivate the investment in additional work to better understand why such apparent discrepancies appear. Some of the FOT and more naturalistic observational work that we believe is needed to better understand these issues is beyond the scope of what individual OEMs and Tier I equipment suppliers can realistically be expected to carry out on their own. We are hopeful that one potential outcome of this project is the encouragement of further collaborative efforts by various stakeholders to undertake additional work in this area. In summary, a number of points should be kept in mind when reviewing this report, and they are elaborated on below. In addition, a number of comments and critiques from industry reviewers and various panel members are also recognized here:

- As discussed throughout this report, with the exception of ESC, the primary limitation in rating the technologies considered here is the relatively limited data available on how each technology performs under real-world driving conditions. While most of the technologies appear to have significant safety potential, for the most part, we know relatively little about how drivers interact with many of these technologies on a daily basis and how that may influence their ultimate performance. More work is needed in this area.
- Following the original submission of this report to AAAFTS, the University of Michigan Transportation Research Institute (Blower, 2014) released a literature review covering some of the same technologies considered here (e.g. ESC, FCW, FCM, and LDW). Generally consistent with our conclusions, the report states that the technologies were "estimated to be substantially effective in reducing their target crash types." However, it was noted that most studies relied on simulation or limited field operational tests. Other than for ESC, it was stated that, "available crash data cannot yet support evaluation of the actual crash experience of the technologies." While we believe that some of the insurance based claim data on crashes and property damage reviewed in our report show an advance in the state of

knowledge somewhat beyond what was considered in the UMTRI report, our substantive views are similar.

- In line with the statements above, representatives of several OEMs expressed their opinion that the data at this point does not seem robust enough to support "even a general rating" of most of these technologies. However, they also emphasized that the information brought together in this report on the technologies and the status of the data is useful. Other representatives commented that the overall ratings appeared reasonable given the data currently available.
    - Noting the exception of ESC, one representative passed on the recommendation that, in terms of presentation to consumers, it might be better to wait until further data is collected to reflect accurate safety benefits of these advanced technologies.
    - One OEM expanded on this by stating: "The highest priority should be educating consumers (before rating) about functionalities of safety technologies to avoid over reliance and/or misunderstanding as well as to improve acceptance."
    - As seen in the case of systems that combine elements of ACC, FCW, and FCM/Autobraking, integrated safety systems are becoming the trend in the industry. It was suggested that it is more desirable to evaluate comprehensive safety rather than individual safety technologies.
- Representatives of an industry group commented that the work to date produced by the project represented a good start for further educating consumers on the different technologies.
    - It was suggested that there might be some value in expanding the adaptive headlight review to include automatic high beam control in addition to steerable headlights. This would allow expansion of the scenario considered from "driving around curves at night" to "night driving."
- The initial ratings provided as part of this proposed rating system represent a "snap-shot" in time and are in some instances based upon "dated" data that represents the best available information. As more data ideally becomes available, it may justify a change in the rating of that technology. Consequently, any rating system should ideally be viewed as dynamic and updated on a periodic basis to provide the most meaningful representation of a system's benefit.
- Just as more data on system performance accumulates over time, technologies continue to be refined and improved, such that a technology class that is given a modest rating today may offer a significantly enhanced safety benefit in next year's model. This also argues for periodic updating of any rating system.
- It is very important to recognize that different implementations of a class of technology may vary widely in their overall effectiveness and the specific scenarios for which they are optimized. Some of this information is noted in the ***Technology Reviews*** in Appendix C and consumer oriented support material drafted as part of this project. Nonetheless, this highlights one of the challenges in providing broad guidance on the potential of a technology class versus providing ratings of specific vehicle models.
    - A representative of one OEM specifically advised that they felt it was most appropriate for the project to be considering generic technologies as a class type as opposed to attempting to evaluate how each manufacturer implemented a technology.

- Another OEM representative noted as a challenge for the rating system the example that ACC based on a single radar would operate over a smaller speed range than ACC based on multiple sensors (radars and/or cameras) that operate over the full speed range of the vehicle. As the system stands now, preference is given to data representing "best in class" performance.
    - This highlights the importance of providing supporting information along with the upper level ratings that makes clear that such implementation differences can exist.

- The current rating approach weights all event types equivalently. One consequence of this is the relatively modest ranking assigned to back-up cameras in terms of *Overall Safety Benefit*. This is a result of the relatively low number of backup event related injuries and fatalities relative to the total number of driving related injuries and fatalities. At the same time, it is recognized that there is a particularly high emotional cost associated with this type of event. NHTSA (2014) estimates that 31 percent of all backup event fatalities involve children under five years of age and another 26 percent are adults 70 years and older; these events often involve family members or other close associations. Societal pressure to do something about such events contributed to NHTSA issuing a final rule on March 31, 2014 mandating rear visibility technology in all new vehicles under 10,000 pounds by May 2018. This type of value assignment is most likely best done outside of the proposed rating system and left to the judgment of societal and/or personal evaluation of the personal significance of a particular type of scenario. This is one of the reasons for incorporating both overall benefit and scenario specific objective ratings in the rating system.
- The question was raised by industry representatives as to whether ESC and back-up cameras should be included in the review as ESC is mandated for all vehicles produced from 2012 forward and rear-view visualization was recently (March 31, 2014) mandated for model years 2018 forward.
  - ESC was specifically included in the initial rating system for scaling purposes as a reference point for a technology that clearly qualified for a top level rating. Keeping ESC in the ratings might have value in encouraging owners of older cars without ESC to consider "upgrading" to derive a clear safety value. It has also been suggested that this might help inform used car shoppers to only consider older cars that include the technology. Nonetheless, it is a reasonable point to be reviewed as to whether inclusion of ESC should be carried forward.
  - As noted, the mandate for back-up visualization technology was issued after this report was essentially completed. Nonetheless, including back-up cameras in the current ratings does provide heuristic value in several areas. For one, it highlights the distinction between purely statistical ratings of significance in terms of absolute numbers of events versus the apparent societal / emotional rating of a particular class of adverse scenario. Furthermore, the observations that the estimated scenario specific benefit for the technology class is currently less than 5 stars and that the currently demonstrated benefit of back-up camera systems is less than the theoretical potential, both highlight that there seem to be gaps in the full realization of this technological concept. This should encourage a closer look at how this class of technology might be further improved. In other words, the fact that a

mandate has now been issued to implement technology to address back-up events does not mean that the push to solve this problem should necessarily be considered over. Further improvement in existing technology implementations appears to be needed, and including back-up cameras in the rating system may have utility in contributing to the discussion around what remains to be done.

- A concern was expressed by one OEM that the project report did not provide a complete description of the database search and the search commands used in the review, and suggested that this limited the ability to update the ratings as additional real-world data is collected. While we believe that the search employed to develop the current review was quite extensive and well supplemented by requests to experts from a wide range of disciplines, as well as OEM and Tier One representatives, we agree that systematic documentation of search terms and databases employed would be useful in any further development of the project. This firm also stated that they understood "the limitation in publicly available data to show actual benefits or potential benefits for specific technologies and in particular the difficulties when comparing systems of different designs and performance characteristics in addressing a specific scenario. However, studies do exist and the MIT team has been successful in gathering extensive information on the technologies in question…"
- In early conceptions of the rating system, we proposed including a range factor within the rating of each technology class to indicate the extent to which variation in effective benefit was present. This was dropped from the proposal put forth here for two reasons. First, a number of members of the Advisory Panel and others consulted on the project expressed the strong opinion that this presented too much detail in the upper-level rating information and was likely to confuse consumers. Second, a number of individuals also strongly argued for the upper-level ratings to reflect the safety benefit of "best in class" systems. The intent here is to err on the side of encouraging serious consideration of technologies with the potential to increase safety.
- One automotive manufacturer expressed concern that a more mature technology may show an apparent advantage in the ratings over a newer technology, as data on the newer technology will initially be limited. The commenter further observed that given the time lag in studies and results on deployed systems, this might have a negative effect in terms of promotion and acceptance, rather than helping to improve customer adoption rate of the technology. This does seem to be a logical concern. One of the reasons for including a rating of theoretical benefit in the system is to provide a best-case estimate of potential while deployment data is sparse. Interestingly enough, one of the newest technologies, forward collision braking, does come out with a higher currently documented benefit rating than all of the other, more established technologies except for ESC. However, this higher demonstrated benefit rating is traceable to a single recent study that was not included in the first draft ratings. This highlights the importance of early studies of real-world behavior to advance confidence in new technologies.
- A number of contributors/reviewers have continued to express concern over our proposed dual rating of estimated safety benefit vs. observed demonstrated benefit in real-world conditions. There seems to be general agreement around the validity and importance of the distinction, but concern that this may be confusing for some consumers. The authors feel strongly that this distinction is important for a number

of reasons, many of which are enumerated in this report, but take the issue of possible confusion quite seriously. It is likely that the proposed system would benefit from further creative input on possible alternative methods of presenting this information to the general public. Along the same lines, focus group testing is also suggested to assess the extent to which a generic technology rating (the focus of this effort) contributes to the education of consumers on the range of capabilities of technologies that they may wish to purchase.

It is clear that distilling the assessment of the potential benefit a particular technology down to a meaningful and appropriate single rating value is challenging and, in some ways, questionable. The proposed rating system represents an attempt to take into account, at a minimum, a number of important concepts such as percentage reduction across all crash events vs. percentage reduction with the specific scenario that a given technology was developed to address. Nonetheless, we believe that the concepts and information developed and drawn together so far over the course of this project make a constructive contribution to efforts to better understand the status of these technologies. At a minimum, the initiation of this project by AAAFTS has stimulated significant discussion and constructive exchange between a broad cross-section of stakeholders concerned with driving safety.

Next Steps

A separate document, *Evaluating Technologies Relevant to the Enhancement of Driver Safety: A Vision beyond Phase I*, was prepared at the request of AAAFTS staff concerning our thoughts regarding possible next steps in the project.

As detailed there, the project was initiated with the vision that it would extend beyond the initial phase of rating a minimum of five technologies. As originally conceived, the intent was to continue adding technologies to the matrix to cover a wider range of relevant technologies. If this rating process and the information and materials developed prove useful, the next logical step would be to periodically update the rating of technologies as they evolve and as improved data becomes available. Two possible approaches to continuing this vision were proposed. The more extensive recommendation suggested that the project be continued to add additional safety technologies to the existing ratings. A possible second round of ratings might consider:

- Blind spot detection
- Lane departure mitigation / lane keeping assist
- Fatigue detection
- Back-up proximity detection sensors
- Pedestrian collision mitigation systems

Collecting relevant data on additional technologies, such as those listed above, and applying the obtained findings to the rating system would have both intrinsic informational value, and would contribute to better evaluating whether the current scaling levels serve their intended purpose or should be adjusted.

Combined References

The following is a combined listing of references cited in the main body of the report and the accompanying appendices.

AAA-FTS. (2008a). Use of Advanced In-Vehicle Technology by Young and Older Early Adopters. Washington, DC. 73.

AAA-FTS. (2008b). Use of Advanced In-Vehicle Technology by Young and Older Early Adopters: Survey Results on Adaptive Cruise Control Systems. In AAA Foundation for Traffic Safety (Ed.), (pp. 2). Washington, DC.

Anderson, R. W. G., Doecke, S. D., Mackenzie, J. R., Ponte, G., Paine, D., & Paine, M (2012). Potential benefits of forward collision avoidance technology. Australia, Centre for Automotive Safety Research (Report CASR106).

Anderson, R. W. G., Hutchinson, T. P., Linke, B.J., & Ponte, G. (2011). Analysis of crash data to estimate the benefits of emerging vehicle technology (CASR094). Adelaide, Australia, Centre for Automotive Safety Research. The University of Adelaide.

Anthikkat, A. P., Page, A., & Barker, R. (2013). Low-speed vehicle run over fatalities in Australian children aged 0-5 years. Journal of Pediatrics and Child Health, 49(5), 388-393.

Ayres, T., Li, L., Trachtmanc, D., & Young, D. (2005). Passenger-side rear-view mirrors: driver behavior and safety. International Journal of Industrial Ergonomics, 35, 157-162.

Balint, A., Fagerlind, H., & Kullgren, A. (2013). A test-based method for the assessment of pre-crash warning and braking systems. *Accident Analysis and Prevention, 59,* 192-199.

Bao, S., LeBlanc, D. J., Sayer, J. R., & Flannagan, C. (2012). Heavy-truck drivers' following behavior with intervention of an integrated, in-vehicle crash warning system: a field evaluation. *Human Factors, 54(5),* 687-697.

Ben-Yaacov, A., Maltz, M., & Shinar, D. (2002). Effects of an in-vehicle collision avoidance warning system on short- and long-term driving performance. *Human Factors, 44(2),* 335-342.

Blower, D. (2014). Assessment of the effectiveness of advanced collision avoidance technologies (Report UMTRI-2014-3). University of Michigan Transportation Research Institute, Ann Arbor, MI.

Brackstone, M., & McDonald, M. (2007). Driver headway: How close is too close on a motorway? *Ergonomics, 50(8),* 1183-1195.

Braitman, K. A., McCartt, A. T., Zuby, D. S., & Singer, J. (2010). Volvo and Infiniti drivers' experiences with select crash avoidance technologies. *Traffic Injury Prevention, 11(3),* 270-278.

Chien, S., Li, L., & Chen, Y. P. (2012). A new braking and warning scoring system for vehicle forward collision imminent braking systems. Paper presented at the 15th International IEEE Conference on Intelligent Transportation Systems, Anchorage, Alaska, USA.

Chouinard, A., & Lecuyer, J. F. (2011). A study of the effectiveness of Electronic Stability Control in Canada. *Accident Analysis and Prevention, 43*(1), 451-460.

Coelingh, E., A. Eidehall, Bengtsson, M. (2010). Collision Warning with Full Auto Brake and Pedestrian Detection -- a practical example of Automatic Emergency Braking. 13th International IEEE Annual Conference on Intelligent Transportation Systems, Madeira Island, Portugal.

Coelingh, E., L. Jakobsson, L., Lind, H., & Lindman, M. (2007). Collision Warning with Auto Brake - A Real-Life Safety Perspective. 20th International Technical Conference on the Enhanced Safety of Vehicles (ESV), Lyon, France.

Consumer Report. (2012). Best and worst rear blind zones. Retrieved August 4, 2012, from http://www.consumerreports.org/cro/2012/03/the-danger-of-blind-zones/index.htm

Dang, J. N. (2004). Preliminary results analysing the effectiveness of electronic stability control (ESC) systems Evaluation Note. National Highway Traffic Safety Administration Washington, D.C. *DOT HS 809 790*.

Eichelberger, A. H., & McCartt, A. T. (2012). Volvo Drivers' Experiences with Advanced Crash Avoidance and Related Technologies. Insurance Institute for Highway Safety. Arlington, VA.

Erke, A. (2008). Effects of electronic stability control (ESC) on accidents: a review of empirical evidence. *Accident Analysis and Prevention, 40*(1), 167-173.

Farmer, C. (2004). Effect of electronic stability control on automobile crash risk. *Traffic Injury Prevention, 5*(4), 317-325.

Farmer, C. M. (2010). Effects of Electronic Stability Control on Fatal Crash Risk. Insurance Institute for Highway Safety. Arlington, VA.

Ferguson, S. A. (2007). The effectiveness of electronic stability control in reducing real-world crashes: a literature review. *Traffic Injury Prevention, 8*(4), 329-338.

Georgi, A., Zimmermann, M., Lich, T., Blank, L., Kickler, N., & Marchthaler, R. (2009). New approach of accident benefit analysis for rear end collision avoidance and mitigation systems. Proceedings of the 21st International Technical Conference on the Enhanced Safety of Vehicles, Stuttgart, Germany.

Green, P. E., & Woodrooffe, J. (2006). The estimated reduction in the odds of loss-of-control type crashes for sport utility vehicles equipped with electronic stability control. *J Safety Res, 37*(5), 493-499.

Griffin, B., Watt, K., Wallis, B., Shields, L., & Kimble, R. (2011). Paediatric low speed vehicle run-over fatalities in Queensland. Injury Prevention, 17(Supplement), i10-13.

Gordon, T., Sardar, H., Blower, D., Ljung Aust, M., Bareket, Z., Barnes, M., et al. (2010). Advanced Crash Avoidance Technologies (ACAT) Program – Final Report of the Volvo-Ford-UMTRI Project: Safety Impact Methodology for Lane Departure Warning – Method Development and Estimation of Benefits. Washington, DC: 218.

Hammond, C., & Wade, M. G. (2005). Forward looking blindspots: A report of A-Pillar induced field-of-view obstruction and driver performance in a simulated rural environment. *Advances in Transportation Studies: An International Journal, Section B5*, 69-81.

Hetrick, S. (1997). Examination of driver lane change behavior and the potential effectiveness of warning onset rules for lane change or "side" crash avoidance systems. (M.Sc.), Virginia Polytechnic Institute and State University, Blacksburg, VA.

HLDI. (2011). Mazda collision avoidance features: initial results. Highway Loss Data Institute. *28(13).*

HLDI. (2012). Mercedes-Benz collision avoidance features: initial results. Highway Loss Data Institute. *29*(7).

HLDI. (2014).Honda Accord collision avoidance features: initial results. Highway Loss Data Institute. *31(2).*

Hoedemaeker, M., & Brookhuis, K. (1998). Behavioural adaptation to driving with an adaptive cruise control (ACC). *Transportation Research Part F : Traffic Psychology and Behaviour, 1*, 95-106.

Hoye, A. (2011). The effects of electronic stability control (ESC) on crashes--an update. *Accident Analysis and Prevention, 43*(3), 1148-1159.
Hurwitz, D., Pradhan, A. K., Fisher, D., Knodler, M., Muttart, J., Menon, R., & Meissner, U. (2010). Backing collisions: a study of drivers' eye and backing behavior using combined rear-view camera and sensor systems. *Injury Prevention, 16*, 79-84.
IHS iSuppli's. (2013, May 16). Cameras in Motor Vehicles to Grow More Than Fivefold by 2020. Retrieved December 6, 2013, from http://www.isuppli.com/Automotive-Infotainment-and-Telematics/MarketWatch/Pages/Cameras-in-Motor-Vehicles-to-Grow-More-Than-Five-fold-by-2020.aspx
IIHS. (2011). Comment to the National Highway Traffic Safety Administration concerning proposed amendments to rearview mirrors safety standard (Vol. NHTSA-2010-0162). Arlington, VA.
IIHS. (2012). They're working: insurance claims data show which new technologies are preventing crashes. *Status Report, 47*(5), 1-7.
IIHS. (2012). http://www.iihs.org/iihs/topics/t/crash-avoidance-technologies/qanda#electronic-stability-control. Retrieved September 15, 2013
IIHS. (2013, September 27, 2013). First crash avoidance ratings under new test program: 7 midsize vehicles earn top marks. Retrieved October 28, 2013, from http://www.iihs.org/iihs/sr/statusreport/article/48/7/1
Isaksson-Hellman, I. & Lindman, M. (2012). The Effect of a Low-Speed Automatic Brake System Estimated from Real Life Data. 56th AAAM Annual Conference Annals of Advances in Automotive Medicine.
Jenness, J. W., Lerner, N. D., Mazor, S. D., Osberg, J. S., & Tefft, B. C. (2007). Use of Advanced In-Vehicle Technology By Young and Older Early Adopters: Results on Sensor-Based Backing Systems and Rear-View Video Cameras. National Highway Transportation Safety Administration. Washington, DC. *DOT-HS-810-828rev*, 178.
Jenness, J. W., Lerner, N. D., Mazor, S., Osberg, J. S., & Tefft, B. C. (2008). Use of advanced in-vehicle technology by young and older early adopters: Survey results on adaptive cruise control systems. National Highway Traffic Safety Administration. Washington, DC. *DOT HS 810 917*, 105.
Jermakian, J. S. (2011). Crash avoidance potential of four passenger vehicle technologies. *Accident Analysis and Prevention, 43*(3), 732-740.
Kidd, D. G. & Brethwaite, A. (2013). Visibility of children behind 2010-13 model year passenger vehicles using glances, mirrors, and backup cameras and parking sensors. Insurance Institute for Highway Safety, Arlington, VA.
Kidd, D.G., Hagoski, B.K., Tucker, T.G., & Chiang, D.P. (2014). Effects of a rearview camera, parking sensor system, and the technologies combined on preventing a collision with an unexpected stationary or moving object. Insurance Institute for Highway Safety, Arlington, VA.
Koziol, J., Inman, V., Carter, M., Hitz, J., Najm, W., Chen, S., Lam, A., Penic, M., Jensen, M., Baker, M., Robinson, M., & Goodspeed, C. (1999). Evaluation of the Intelligent Cruise Control System Volume I – Study Results. National Highway Traffic Safety Administration. Washington, D.C. *DOT HS 808 969*, 310.
Kusano, K. D. & Gabler, H.C. (2012). Safety Benefits of Forward Collision Warning, Brake Assist, and Autonomous Braking Systems in Rear-End Collisions. *IEEE Transactions on Intelligent Transportation Systems, 13(4),* 1546-1555.
LeBlanc, D., Sayer, J., Winkler, C., Ervin, R., Bogard, S., Devonshire, J., Mefford, M., Hagan, M. T., Bareket, Z., Goodsell, R., & Gordon, T. (2006). Road Departure Crash

Warning System Field Operational Test: Methodology and Results. University of Michigan Transportation Research Institute. Ann Arbor, MI. (UMTRI-2006-9-1).
Larsson, A. F. (2012). Driver usage and understanding of adaptive cruise control. *Applied Ergonomics, 43(3),* 501-506.
Lie, A. (2012). Nonconformities in real-world fatal crashes--electronic stability control and seat belt reminders. Traffic *Injury Prevention*, 13(3), 308-314.
Lie, A., Tingvall, C., Krafft, M., & Kullgren, A. (2006). The effectiveness of electronic stability control (ESC) in reducing real life crashes and injuries. *Traffic Injury Prevention*, *7(1),* 38-43.
Lee, J. D., McGehee, D. V., Brown, T. L., & Reyes, M. L. (2002). Collision warning timing, driver distraction, and driver response to imminent rear-end collisions in a high-fidelity driving simulator. *Human Factors, 44*(2), 314-334.
Llaneras, R. E. (2006). Exploratory study of early adopter, safety-related driving with advanced technologies. for NHTSA by Westat. *DOT HS 809 972*.
Llaneras, R. E., Neurauter, M. L., & Green, C. A. (2011). *Factors moderating the effectiveness of rear vision systems: what performance-shaping factors contribute to unexpected in-path obstacles when backing?* Paper presented at the SAE 2011 World Congress & Exhibition, Warrendale, PA.
Lund, A. (2013). *Drivers and Driver Assistance Systems: How well do they match*? Presentation at the 7th International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design, Bolton Landing, NY, USA, June 18, 2013.
Lund, A. (2014). Personal communication expanding upon interpretation of data presented in Lund (2013) and other IIHS reports; April 3, 2014.
Markvollrath, Schleicher, S., & Gelau, C. (2011). The influence of cruise control and adaptive cruise control on driving behaviour--a driving simulator study. *Accident Analysis and Prevention, 43(3*), 1134-1139.
Marsden, G., McDonald, M., & Brackstone, M. (2001). Towards an understanding of adaptive cruise control. *Transportation Research Part C: Emerging Technologies, 9*, 33-51.
Matsui, Y., Han, Y., & Mizuno, K. (2011). Performance of collision damage mitigation braking systems and their effects on human injury in the event of car-to-pedestrian accidents. *Stapp Car Crash J, 55,* 461-478.
Mazzae, E. N. (2008). On-Road Study of Drivers' Use of Rearview Video Systems (ORSDURVS). National Highway Traffic Safety Administration. Washington, D.C. *No. DOT HS 811 024*, 140.
Mazzae, E. N. (2010). Drivers' Use of Rearview Video and Sensor-Based Backing Aid Systems in a Non-Laboratory Setting. National Highway Traffic Safety Administration. Washington, D.C. *No. NHTSA-2010-0162*, 16.
Mazzae, E. N. (2013). Rearview Video System Use by Drivers of a Sedan in an Unexpected Obstacle Event National Highway Traffic Safety Administration. Washington, D.C. *No. NHTSA-2010-0162*, 39.
Mazzae, E. N., & Barickman, F. (2009). Direct rear visibility of passenger cars: Laser-based measurement development and findings for late model vehicles. National Highway Traffic Safety Administration. Washington, DC. *DOT HS-811-174*.
Mazzae, E. N., & Garrott, W. R. (2008). Light vehicle rear visibility assessment. National Highway Traffic Safety Administration. Washington, DC. *DOT HS-810-909.*
McLaughlin, S., Hankey, J., Green, C. A., & Larsen, M. (2004a). Discomfort Glare Ratings of Swiveling HID Headlamps. SAE International. Warrendale, PA. *2004-01-2257*.

McLaughlin, S., Hankey, J., Green, C. A., & Larsen, M. (2004b). Target Detection Distances and Driver Performance with Swiveling HID Headlamps. *SAE TECHNICAL Paper Series, 2004-01-2258*.

Najm, W. G., Stearns, M. D., Howarth, H., Koopmann, J., & Hitz, J. (2006). Evaluation of an automotive rear-end collision avoidance system. National Highway Traffic Safety Administration. Washington, DC. (DOT HS-810-569).

Navarro, J., Mars, F., Forzy, J. F., El-Jaafari, M., & Hoc, J. M. (2010). Objective and subjective evaluation of motor priming and warning systems applied to lateral control assistance. *Accident Analysis and Prevention, 42(3),* 904-912.

Neeman, T., Wylie, J., Attewell, R., Glase, K., & Wallace, A. (2002). Driveway Deaths: Fatalities of Young Children in Australia as a Result of Low-Speed Motor Vehicle Impacts (Vol. Road Safety Report no. CR208). Canberra, Australia: Australian Transport Safety Bureau.

Nhan, C., Rothman, L., Slater, M., & Howard, A. (2009). Back-over collisions in child pedestrians from the Canadian hospitals injury reporting and prevention program. Traffic and Injury Prevention, 10(4), 350-353.

NHTSA. (2006). Vehicle Backover Avoidance Technology Study. U.S. Department of Transportation. Washington, DC. 65.

NHTSA. (2007). *Final rule. Docket no. NHTSA-2007-27662; 49 CFR Parts 571 and 585 - Federal Motor Vehicle Safety Standards, Electronic stability control systems, Controls and displays*. Washington, DC: Office of the Federal Register.

NHTSA. (2008). Fatalities and injuries in motor vehicle backing crashes. U.S. Department of Transportation. Washington, DC. *DOT HS 811 144.*

NHTSA. (2010). Federal Motor Vehicle Safety Standard, Rearview mirrors; Federal Motor Vehicle Safety Standard, Low-Speed Vehicles Phase-In Reporting Requirements; proposed rule (Vol. 75(234)).

NHTSA. (2013). Releases Policy on Automated Vehicle Development, Preliminary Statement of Policy Concerning Automated Vehicles. U.S. Department of Transportation.

NHTSA. (2014). Federal Motor Vehicle Safety Standards; Rear Visibility; Final Rule. U.S. Department of Transportation. Federal Register, 79(66),

Nodine, E. E., Lam, A.H., Najm, W.G., & Ference, J.J. (2011). Safety impact of an integrated crash warning system based on field test data. 22nd International Technical Conference on the Enhanced Safety of Vehicles. Washington, DC.

Page, Y., Foret-Bruno, J., & Cuny, S. (2005). Are expected and observed effectiveness of emergency brake assist in preventing road injury accidents consistent? Paper presented at the 19th International Technical Conference on Enhanced Safety of Vehicles, Washington, D.C.

Peng, Y., Boyle, L. N., & Hallmark, S. L. (2013). Driver's lane keeping ability with eyes off road: Insights from a naturalistic study. *Accident Analysis and Prevention, 50,* 628-634.

Piccinini, G. F., Simoes, A., Rodrigues, C. M., & Leitao, M. (2012). Assessing driver's mental representation of Adaptive Cruise Control (ACC) and its possible effects on behavioural adaptations. *Work, 41 Suppl 1,* 4396-4401.

Pinkney, K., Smith, A., Mann, N., Mower, G., Davis, A., & Dean, J. (2006). Risk of pediatric back-over injuries in residential driveways by vehicle type. *Pediatric Emergency Care, 6*, 402-407.

Roberts, I., Norton, R., & Jackson, R. (1995). Driveway-related child pedestrian injuries: a case-control study. *Pediatrics in Review, 95,* 405-408.

Robinson, B., Hulshof, W., Cookson, R., Cuerden, R., Hutchins, R., & Delmonte, (2011). Cost benefit evaluation of advanced primary safety systems. Transport Research Laboratory, Report No. PPR586.

Robertson, R. D., Vanlaar, W. G. M., Marcoux, K., & McAteer, H. J. (2012). Vehicle safety features: knowledge, perceptions, and driving habits. Traffic Injury Research Foundation. 77.

Rosen, E., Kallhammer, J. E., Eriksson, D., Nentwich, M., Fredriksson, R., & Smith, K. (2010). Pedestrian injury mitigation by autonomous braking. *Accident Analysis and Prevention, 42(6),* 1949-1957.

Rudin-Brown, C. M., & Burns, P. C. (2007, June 3-6). The secret of electronic stability control (ESC). Paper presented at the Canadian Multidisciplinary Road Safety Conference XVII, Montreal, QC.

Rudin-Brown, C. M., Burns, P. C., Jenkins, R., Whitehead, T., & LeBlond, O. (2011). ESC (Electronic Stability Control) Public and Driver Surveys (Vol. TP 14848E). Ottawa, ON: Transport Canada.

Rudin-Brown, C. M., & Parker, H. A. (2004). Behavioural adaptation to adaptive cruise control (ACC): implications for preventive strategies. *Transportation Research Part F : Traffic Psychology and Behaviour, 7,* 59-76.

Sayer, J. R., Bogard, S. E., Buonarosa, M. L., LeBlance, D. J., Funkhouser, D. S., Bao, S., Blankespoor, A. D., & Winkler, C. B. (2011). Integrated vehicle-based safety systems light-vehicle field operational test key findings report. National Highway Traffic Safety Administration. Washington, DC.

Sayer, J. R., Bogard, S. E., Funkhouser, D., LeBlanc, D. J., Bao, S., Blankespoor, A. D., Buonarosa, M. L., & Winkler, C. B. (2010). Integrated vehicle-based safety systems heavy-truck field operational test key findings report. National Highway Traffic Safety Administration. Washington, DC. *DOT HS-811-362*.

Sivak, M., Flannagan, M. J., Traube, E. C., Aoki, M., & Sayer, J. (1994). Evaluation of an active headlight system. University of Michigan Transportation Research Institute. Ann Arbor, MI. *Report no. UMTRI-94-17*.

Sivinski, R. (2011). Crash Prevention Effectiveness of Light-Vehicle Electronic Stability Control: An Update of the 2007 NHTSA Evaluation. NHTSA. Washington, DC. 29.

Strandroth, J., Rizzi, M., Olai, M., Lie, A., & Tingvall, C. (2012). The effects of studded tires on fatal crashes with passenger cars and the benefits of electronic stability control (ESC) in Swedish winter driving. Accident Analysis and Prevention, 45, 50-60.

Stanton, N. A., & Young, M. S. (2005). Driver behaviour with adaptive cruise control. *Ergonomics, 48(10),* 1294-1313.

Sullivan, J. M., Flannagan, M. J., & Schoettle, B. (2002). The Appearance of Bending Light From Other Vehicles. The University of Michigan Transportation Research Institute. *UMTRI-2002-2*, 28.

Tanaka, S., Mochida, T., Aga, M., & Tajima, J. (2012). Benefit estimation of a lane departure warning system using ASSTREET. *SAE International Journal of Passenger Cars – Electronic and Electrical Systems, 5(1),* 133-145.

Tapia-Espinoza, R., & Torres-Torriti, M. (2013). Robust lane sensing and departure warning under shadows and occlusions. Sensors (Basel), 13(3), 3270-3298.

Terry, H. R., Charlton, S. G., & Perrone, J. A. (2008). The role of looming and attention capture in drivers' braking responses. *Accident Analysis and Prevention, 40(4),* 1375-1382.

Thatcham. (2013). All ESC are created equal, but some are more equal than others? *Thatcham Research News, Special Edition, 1,* 13.

Tijerina, L., Blommer, M., Curry, R., Greenberg, J., Kochhar, D., Simonds, C., & Watson, D. (2010). Effects of adaptive lane departure warning system on driver response to a surprise event. *Transportation Research Record, 2185(2185),* 1-7.
TIRF. (2013). http://brainonboard.ca/. Retrieved 2013-09-16
Transport Canada. (2011, 2011-06-20). http://www.tc.gc.ca/eng/roadsafety/safevehicles-1186.htm. Retrieved 2013-09-16
USDOT FMCSA. (2013, October 28, 2013). Standard No. 126; Electronic stability control systems. Retrieved December 8, 2013, from http://www.fmcsa.dot.gov/rules-regulations/administration/fmcsr/fmcsrruletext.aspx?reg=571.126
U.S. House of Representatives. (2007). Cameron Gulbransen Kids Transportation Safety Act of 2007. Washington, DC: US Congress.
Van Auken, R. M., Zellner, J.W., Chiang, D.P., Kelly, J., Silberling, J.Y., Dai, R., Broen, P.C., Kirsch, A.M., & Sugimoto, Y. (2011). Advanced Crash Avoidance Technologies (ACAT) Program - Final Report of the Honda-DRI Team, Volume I: Executive Summary and Technical Report. Washington, DC, US DOT NHTSA: 285.
Wege, C., Will, S., & Victor, T. (2013). Eye movement and brake reactions to real world brake-capacity forward collision warnings-A naturalistic driving study. *Accident Analysis and Prevention, 58*, 259-270.
Wilson, B. H., Stearns, M. D., Koopmann, J., & Yang, C. Y. (2007). Evaluation of a road-departure crash warning system. National Highway Traffic Safety Administration. Washington, DC. (DOT-HS-810-854).
Xiong, H., Boyle, L. N., Moeckli, J., Dow, B. R., & Brown, T. L. (2012). Use patterns among early adopters of adaptive cruise control. *Human Factors, 54*(5), 722-733.
Yasuda, H., A. Kozato, et al. (2011). A forward collision warning (FCW) performance evaluation. Proceedings of the 22nd International Technical Conference on the Enhanced Safety of Vehicles. Washington, DC.
Young, M. S., & Stanton, N. A. (2004). Taking the load off: investigations of how adaptive cruise control affects mental workload. *Ergonomics, 47(9),* 1014-1035.
Zador, P., Stein, H. S., Wright, P., & Hall, J. (1987). Effects of chevrons, post-mounted delineators, and raised pavement markers on driver behaviour at roadway curves. *Transportation Research Record, 11114*, 1-10.

Appendix A: Rational Scaling Structure for Level Scaling

The table below presents the scaling structure used for the current rating levels used in the proposed system. See main text for details.

Actual Scenario	min.	max.	range	scaling
Level 1	1	10	9	
Level 2	11	22	11	2
Level 3	23	38	15	4
Level 4	39	60	21	6
Level 5	61+			

Actual Overall	min.	max.	range	scaling
Level 1	1	5	4	
Level 2	6	11	5	1
Level 3	12	19	7	2
Level 4	20	30	10	3
Level 5	31+			

Theory Scenario	min.	max.	range	scaling
Level 1	1	12	11	
Level 2	13	28	15	4
Level 3	29	50	21	6
Level 4	51	80	29	8
Level 5	81+			

Theory Overall	min.	max.	range	scaling
Level 1	1	6	5	
Level 2	7	14	7	2
Level 3	15	25	10	3
Level 4	26	40	14	4
Level 5	41+			

Appendix B: Advisory Panel

A panel of experts was convened to provide critique and input on all aspects of the rating system and data being considered for rating purposes. The group initially assembled in Washington, DC on July 19, 2013 and continued discussions over five conference calls (October, November and December 2013, and March 2014). Numerous panel members provided directed comments via personal emails and conversations independent from the panel meetings. Input from this group provided pivotal feedback on various revisions to the rating system, review of the overall methodology selected for rating and detailed feedback on the Technology Review sheets assembled for each technology.

The current materials were developed taking into account feedback from panel members, but the content and interpretation should not be seen as necessarily representative of any individual member's opinion or of organizations with which they are affiliated.

Chair
Joseph Coughlin - Director MIT AgeLab & New England University Transportation Center

Members
Academic Research

- Dan McGehee - Director of the Human Factors and Vehicle Safety Research Division at the University of Iowa Public Policy Center
- Jim Sayer - Research Scientist in the Human Factors Group at the University of Michigan Transportation Research Institute

Automotive Industry

- Mike Cammisa - Director, Safety - Global Automakers
- Scott Schmidt - Senior Director, Safety & Regulatory Affairs - Alliance of Automobile Manufacturers

Government Agency

- Jennifer Dang – Chief of NHTSA's New Car Assessment Program (NCAP)
- Erin Sauber-Schatz - Acting Team Lead/Epidemiologist Transportation Safety Team Division of Unintentional Injury Prevention National Center for Injury Prevention and Control, CDC

Insurance industry

- Adrian Lund - President of the Insurance Institute for Highway Safety and the affiliated Highway Loss Data Institute

Other Specialists

- Joseph Carra - NHTSA Retired

Consumer Safety Advocate

- Paul Santos - Santos Family Foundation

Representatives of the "Consumer"

- Jake Nelson - Director of Traffic Safety Advocacy & Research for AAA
- David Nguyen - Manager, Automotive Engineering AAA National Office

Observers

- Jurek Grabowski - Director of Research AAA Foundation for Traffic Safety
- Peter Kissinger - President & CEO AAA Foundation for Traffic Safety

Appendix C – Technology Review sheets

In their current form, the Technology Review sheets are intended for internal reference. They identify relevant objective data collected on the safety benefit of each of the technologies. Information relevant to factors that may impact effectiveness is also identified. Note that some of the latter material includes listings of concerns or possible issues raised in either the research field or by industry sources. Similarly, industry comments on driver responsibility have been noted in a number of limitations entries. If an entry does not include a source citation, then it should be taken as opinion or hypothesis as opposed to necessarily being data based.

In addition to significant input from a number of OEM representatives, Advisory Panel members, and other advisors already mentioned, we would like to express our appreciation to Richard Young who commented extensively on early versions of several of the technology reviews. However, as noted elsewhere, responsibility for the summarization presented remains with the authors of this report.

Electronic Stability Control (ESC)

Initial Ratings	**Overall**	**Scenario Specific**
Potential Overall Benefit	☆☆☆☆☆	☆☆☆☆☆
Benefit Currently Documented	★★★★★	★★★★★

What is the technology?

- Electronic Stability Control (ESC) is designed to help a driver maintain or regain control of the vehicle in difficult driving situations, such as during unexpected turns or while negotiating icy roads. ESC systems continuously monitor actual vehicle motion (tire movement) and driver's intention (steering wheel activity) to sense a loss of traction or slippage. In such situations, ESC systems apply brakes independently to each wheel to counter oversteer and understeer conditions.
  - Some ESC systems adjust tire suspension and can reduce engine power until control is regained.
- ESC is a level 1 vehicle automation system (Function-Specific Automation) (NHTSA, 2013).

Crash Reduction/Prevention

- An Insurance Institute for Highway Safety (IIHS) report (C. Farmer, 2004) based on an analysis of the Fatality Analysis Reporting System (FARS) for all fatal crashes in the United States over 3 years (2001–2003) found that:
  - ESC was found to have reduced single-vehicle fatal crash involvement risk by 56 percent (C.I. 39–68).
  - This translates to an estimated 34 percent reduction in overall fatal crash involvement risk (C.I. 21–45)(C. Farmer, 2004).
- In a follow-up analysis considering a ten year period, Farmer (C. M. Farmer, 2010) reported somewhat smaller effectiveness values while providing a number of explanations for this decline. The updated summary finds a reduced fatal crash involvement risk of 33%:
  - 20% for multiple-vehicle crashes and 49% for single vehicle crashes.
  - Effectiveness estimates were 30% for cars and 35% for SUV's, although these differences were not statistically significant.
- Using data from the NASS General Estimates System (GES) for Sport utility vehicles, Green and Woodrooffe (2006) showed that the odds of a loss-of-control crash for sport utility vehicles equipped with ESC was reduced by 70.3%.
  - Both genders and all age groups benefited equally from the system.
  - With respect to driver age, the maximum percentage reduction of 73.6% occurred at age 27.
- Fergusson (2007), in a 2003 to 2006 literature review, found that the overwhelming majority of real-world crash studies find ESC to be highly effective in reducing single-

vehicle crashes in cars and SUVs. Single vehicle crash risk was reduced by 33-35% for standard passenger vehicles and 56-67% for SUV's. Additional breakdowns:

  - Fatal single-vehicle crashes involving small cars were reduced by about 30–50% and SUVs by 50–70%.
  - Fatal rollover crashes were estimated to be about 70–90% lower with ESC regardless of vehicle type.
  - A number of studies found improved effectiveness in reducing crashes when road conditions are slippery.
  - ESC does not reduce the overall occurrence of multi-vehicle crashes, but does reduce the number of fatal multi-vehicle crashes by 17-38%.

- Erke (2008) summarized the effects of ESC from a number of studies in a meta-analysis:
  - Large reductions of single vehicle crashes were found (-49%; 95% confidence interval -55% to -42%), and smaller but still statistically significant reductions of head-on collisions (-13%; 95% confidence interval -17% to -8%).
  - Multi-vehicle fatal crashes are also reduced (-32%; 95% confidence interval -43% to -20%).
  - However, the studies vary in their effect size estimates, especially for single vehicle crashes. Results of studies on single vehicle crashes produce larger effect size estimates than are expected based on the total number of crashes that could be affected by ESC, suggesting an upward bias in the 49% single vehicle crash reduction estimate. Unspecified properties of the vehicles, time trends, and driver behavior may have biased the single vehicle effect estimates too high.
- Based on statistical analyses of the Fatality Analysis Reporting System (FARS) and National Automotive Sampling System Crashworthiness Data System (NASS CDS) data from 1997 to 2009, ESC has the potential to prevent 72% of car rollovers and 64% of SUV rollovers that would otherwise occur in single-vehicle crashes (Sivinski, 2011).
- ESC is effective for single-vehicle crashes (18.6% effectiveness across all crash severities, 49.3% effectiveness for injury crashes) (Chouinard & Lecuyer, 2011). The results of the study also show that ESC is effective in Canadian weather conditions (i.e. on ice, snow and slush). The effectiveness of ESC on roads covered with ice, snow and slush is 51.1% across all severities and 71.1% for injury crashes.
- According to Insurance Institute for Highway Safety and the U.S. National Highway Traffic Safety Administration, one-third of fatal collisions could be prevented by the use of the ESC (Dang, 2004; IIHS, 2012).
- ESC has been shown to be effective in different weather conditions (i.e. on ice, snow and slush)(Chouinard & Lecuyer, 2011).

Consumer Awareness & Trust

- Rudin-Brown et al. (2009) conducted two separate telephone surveys evaluating Canadian drivers' perceptions and awareness of ESC. The first surveyed 500 randomly selected owners of passenger vehicles. The second survey contacted 1,017 owners of 2006-2008 ESC-equipped passenger vehicles.
    - Results indicated that awareness of ESC was low. When prompted to identify vehicle safety features, only 1% of the people surveyed mentioned ESC, or a branded equivalent. Out of the first 500 surveyed, sixty percent of drivers had never heard of ESC, and less than 5% were aware that they own a vehicle equipped with ESC.
    - While ESC drivers were much more likely than drivers of other vehicles to be aware of ESC (77% vs. 39%) and whether their own vehicle was ESC-equipped (63% vs. 8%), 23% had never heard of it.
    - Ninety percent of drivers who knew that their vehicle was equipped with ESC believed that ESC had made it safer to drive, and reported being confident that ESC would work in an emergency.
    - Twenty-three percent of ESC owners who knew their vehicle was equipped with ESC reported noticing long-lasting changes in their driving behavior since they began driving the vehicle.
- In a recent survey conducted by Traffic Injury Research Foundation (TIRF), a total of 2,506 Canadians completed a poll on major available safety technologies (832 over the phone and 1,674 online)(Robertson, Vanlaar, Marcoux, & McAteer, 2012).
    - The results showed that only 31.4% of them were familiar with ESC (Female: 20.3%, Male: 44.2%).
    - 41% of the drivers said that ESC could make them a better driver (Female: 37.5%, Male: 44.9%).
    - 65.5% of male drivers perceived easiness of use of ESC whereas female drivers reported a 49.4% of perceived easiness (total: 56.9%).
    - 59.5% of drivers reported they would use ESC in the future (intention to use) (Female: 53.7%, Male: 66.2%).

Mobility Significance

- No substantive research has been identified to date that examined the mobility significance of ESC systems.

Other Benefits

- None identified

Technology Penetration

- In the US, regulation FMVSS No. 126 (Electronic Stability Control Systems Indicative Test for Compliance) requires that all cars be manufactured with ESC technology by 2012. This regulation has been proposed as a GTR (Global Technical Regulation) (National Highway Traffic Safety Administration, 2007).
- The only way to get ESC is to buy a new or used vehicle that is equipped with ESC. It cannot be installed as an add-on package.

Frequency of Use

- Rudin-Brown et al. (2009) conducted two separate telephone surveys evaluating Canadian drivers' perceptions and awareness of ESC. The first surveyed 500 randomly selected owners of passenger vehicles. The second survey contacted 1,017 owners of 2006-2008 ESC-equipped passenger vehicles.
    - ESC is automatically "on" whenever the engine is started. In some models, ESC can be turned off by the driver. If so, a telltale lamp will normally illuminate on the instrument cluster. However, the system will automatically be turned back on at the next ignition. The owner's manual should be consulted to learn how ESC works for a given vehicle.
    - The survey did not report how frequently drivers may have turned off the ESC system.

Training and Education

- No formal studies were identified that examined the impact training/education on drivers' interaction with ESC.
- Consistent with Thatcham's statement (Thatcham, 2013), in order to allow the full intended safety benefits of ESC to reach consumers, vehicle manufacturers are encouraged to market ESC-equipped vehicles in a responsible, safe, and realistic manner.
    - Driver training and safety organizations are also encouraged to provide balanced educational information regarding ESC to their students (Thatcham, 2013).

Behavior Adaptation

- Although past and emerging research indicates that ESC is effective in reducing crash rates and saving lives, and its inclusion in all vehicle platforms is encouraged, it may be speculated that some drivers may develop an over-reliance on ESC that could offset or reduce its overall effectiveness, a phenomenon known as 'behavioral adaptation' (Thatcham, 2013).
    - While potential changes in driver behavior are of concern, ESC's proven effectiveness in reducing the likelihood of being involved in a serious crash outweighs any potential increases in unsafe driving due to behavioral adaptation (C.M. Rudin-Brown, Burns, Jenkins, Whitehead, & LeBlond, 2008).

Auditory Demand

- Some ESC systems utilize a sharp alarm sound, but the effects of different alarm types on the driver and any resulting impact on the effectiveness of ESC have not been studied.

Visual Demand

- Some ESC systems utilize a visual alarm, but the effects of different alarm types on the driver and any resulting impact on the effectiveness of ESC have not been studied.

Haptic Demand

- In this review, no information was found that indicated that haptic warning signals for ESC were used by any vehicle manufacturer.

Cognitive Demand

- The system activates only in situations where the driver is highly likely to be taxed by the demands of maintaining vehicular control. No research has been identified that examines the specific cognitive effects of ESC.
    - It might be hypothesized that ESC engagement could reduce cognitive demand, because it reduces the loss of vehicle control which arguably is a high cognitive demand event. However, this suggestion should be evaluated rather than assumed.

Vehicle Type

- In a meta-analysis conducted by Hoye (2011), ESC was often found to be more effective in Sports Utility Vehicles (SUVs) than in passenger vehicles. This is likely due to many SUV designs having significantly higher centers of gravity than typical passenger vehicles and are, as a consequence, inherently less stable in turning conditions. Such vehicle designs thus show greater percentage improvement with the addition of ESC.
- Since ESC is embedded differently in every vehicle, some sportier models allow more wheel spin and sliding, while still maintaining control. This may influence ESC efficiency, but no studies to date have separated out these effects (Ferguson, 2007).
- On some four-wheel drive vehicles, ESC will turn off when you shift to the low range of four-wheel drive. A dashboard light or message is typically provided to indicate to the driver when this occurs (C.M. Rudin-Brown et al., 2008).

Limitations / Failure Conditions

- ESC systems are not optimized for operation in contact with loose surfaces such as gravel, soft snow, and mud. Some vehicles provide an override switch or other mechanism for disengaging the ESC system if a driver experiences difficulty maneuvering under such conditions.
- This technology does not and cannot change the laws of physics. If a vehicle is traveling too fast for road conditions or is not maintained properly (ex. tires and brakes), an ESC equipped vehicle can still lose control.

Differences between Implementations

- Consumers should be made aware that ESC performance can vary between vehicle models (i.e. lateral displacement of the vehicle, angle maintenance). (Thatcham, 2013).
    - Each ESC system is tuned by the manufacturer to work with the chassis dynamics of each vehicle to provide safety benefits while balancing intentional handling characteristics of a particular brand. Individual model tuning may provide more advantage in one loss of control scenario and less in another.
    - It is also worth noting that vehicle models that crashed more frequently before ESC was introduced are generally the models that have shown the greatest reduction in crashes after ESC was added.

References

Chouinard, A., & Lecuyer, J. F. (2011). A study of the effectiveness of Electronic Stability Control in Canada. *Accident Analysis and Prevention, 43*(1), 451-460.

Dang, J. N. (2004). Preliminary results analyszing the effectiveness of electronic stability control (ESC) systems Evaluation Note. National Highway Traffic Safety Administration Washington, D.C. *DOT HS 809 790*.

Erke, A. (2008). Effects of electronic stability control (ESC) on accidents: a review of empirical evidence. *Accident Analysis and Prevention, 40*(1), 167-173.

Farmer, C. (2004). Effect of electronic stability control on automobile crash risk. *Traffic Inj Prev, 5*(4), 317-325.

Farmer, C. M. (2010). Effects of Electronic Stability Control on Fatal Crash Risk. Insurance Institute for Highway Safety. Arlington, VA.

Ferguson, S. A. (2007). The effectiveness of electronic stability control in reducing real-world crashes: a literature review. *Traffic Inj Prev, 8*(4), 329-338.

Green, P. E., & Woodrooffe, J. (2006). The estimated reduction in the odds of loss-of-control type crashes for sport utility vehicles equipped with electronic stability control. *J Safety Res, 37*(5), 493-499.

Hoye, A. (2011). The effects of electronic stability control (ESC) on crashes--an update. *Accident Analysis and Prevention, 43*(3), 1148-1159.

IIHS. (2012). http://www.iihs.org/iihs/topics/t/crash-avoidance-technologies/qanda#electronic-stability-control. Retrieved September 15, 2013

National Highway Traffic Safety Administration. (2007). *Final rule. Docket no. NHTSA-2007-27662; 49 CFR Parts 571 and 585 - Federal Motor Vehicle Safety Standards, Electronic stability control systems, Controls and displays*. Washington, DC: Office of the Federal Register.

NHTSA. (2013). Releases Policy on Automated Vehicle Development, Preliminary Statement of Policy Concerning Automated Vehicles. U.S. Department of Transportation.

Robertson, R. D., Vanlaar, W. G. M., Marcoux, K., & McAteer, H. J. (2012). Vehicle safety features: knowledge, perceptions, and driving habits. Traffic Injury Research Foundation. 77.

Rudin-Brown, C. M., Burns, P. C., Jenkins, R., Whitehead, T., & LeBlond, O. (2008). ESC (Electronic Stability Control) Public and Driver Surveys. Transport Canada. (TP 14848E).

Rudin-Brown, C. M., Jenkins, R. W., Whitehead, T., & Burns, P. C. (2009). Could ESC (Electronic Stability Control) change the way we drive? *Traffic Inj Prev, 10*(4), 340-347.

Sivinski, R. (2011). Crash Prevention Effectiveness of Light-Vehicle Electronic Stability Control: An Update of the 2007 NHTSA Evaluation. NHTSA. Washington, DC. 29.

Thatcham. (2013). All ESC are created equal, but some are more equal than others? *Thatcham Research news, special edition 1,* 13.

Other Key Sources Reviewed

http://www.fmcsa.dot.gov/rules-regulations/administration/fmcsr/fmcsrruletext.aspx?reg=571.126

Lie, A. (2012). Nonconformities in real-world fatal crashes--electronic stability control and seat belt reminders. Traffic Inj Prev, 13(3), 308-314.

Lie, A., Tingvall, C., Krafft, M., & Kullgren, A. (2006). The effectiveness of electronic stability control (ESC) in reducing real life crashes and injuries. Traffic Inj Prev, 7(1), 38-43.

Rudin-Brown, C. M., & Burns, P. C. (2007, June 3-6). The secret of electronic stability control (ESC). Paper presented at the Canadian Multidisciplinary Road Safety Conference XVII, Montreal, QC.

Rudin-Brown, C. M., Burns, P. C., Jenkins, R., Whitehead, T., & LeBlond, O. (2011). ESC (Electronic Stability Control) Public and Driver Surveys (Vol. TP 14848E). Ottawa, ON: Transport Canada.

Strandroth, J., Rizzi, M., Olai, M., Lie, A., & Tingvall, C. (2012). The effects of studded tires on fatal crashes with passenger cars and the benefits of electronic stability control (ESC) in Swedish winter driving. Accident Analysis and Prevention, 45, 50-60.

USDOT FMCSA. (2013, October 28, 2013). Standard No. 126; Electronic stability control systems. Retrieved December 8, 2013, from http://www.fmcsa.dot.gov/rules-regulations/administration/fmcsr/fmcsrruletext.aspx?reg=571.126

Adaptive (Automatic) Cruise Control (ACC)

Initial Ratings	**Overall**	**Scenario Specific**
Potential Overall Benefit	☆	☆☆
Benefit Currently Documented	★	★★

What is the technology?

- Automatic Cruise Control (ACC) uses distance sensing technology that automatically slows down and speeds up the vehicle to maintain a constant distance between the vehicle and the vehicle directly ahead.
- ACC differs from previous cruise control systems that could only maintain the vehicle's current speed, without taking the speed of other vehicles into account. It still allows a driver to maintain a set speed when no other vehicles are nearby, but if a vehicle is sensed in the forward view by ACC, it can maintain a set time headway to the forward vehicle.
- If the forward vehicle picks up speed, the automatic cruise control system will increase the speed of the vehicle such that the set headway gap is maintained up until the vehicle reaches the programed cruise speed set by the driver. In many vehicles the ACC will not automatically accelerate the vehicle if the vehicle slows below some threshold level.
- Automatic cruise control will not perform emergency braking. ACC will only perform moderate braking. For emergency braking a separate AEB system is required.
- Automatic Cruise Control is a level 1 vehicle automation system (Function-Specific Automation) (NHTSA, 2013).

Crash Reduction/Prevention

- The one publically available report that was identified that provided an estimate of potential safety benefits of ACC was a NHTSA sponsored FOT (Koziol et al., 1999). The authors concluded that if such systems were fully deployed and utilized at the engagement rate seen in the FOT, it was estimated that the number of collisions on freeways for travel velocities above 40 km/h would be reduced by 17% for two specified scenarios. This estimate would correspond to a reduction in the number of police-reported rear-end collisions by about 13,000 in 1996 and this was interpreted as indicating a fairly strong benefit compared to manual driving. However, as a percentage of total crashes of all types, this would correspond to less than 1%.
    - Scenario 1 – when an ACC equipped vehicle approached a slower vehicle traveling at a constant velocity
    - Scenario 2 – when a lead vehicle decelerated in front of an ACC equipped vehicle.
    - It was noted that additional safety benefits would be expected from a reduction in other rear-end collisions involving cut-ins and lane changes and from use of ACC on roadways other than freeways; however, benefit estimates for these scenarios were not examined in the FOT. Drivers were found to engage the system for 6 % of the time on arterials and 11% on state highways.

- A senior staff member at IIHS confirmed in personal communication (12/2013) that they were not aware of any other theoretical estimates of crash prevention or mitigation benefits of ACC or any reports that measured benefits for ACC isolated from other, related safety features. Since more recent IIHS work has considered vehicles that frequently combine ACC with forward collision warning (FCW) and, increasingly, autobrake features, it is seen as difficult to isolate the effects of the component systems.

Consumer Awareness & Trust

- In a recent survey conducted by Eichelberger and McCartt (2012), Volvo drivers were asked whether they would want specific technologies that they currently possess in their next vehicle. Ninety-three percent reported they would want ACC again.
    - Moreover, 49% mentioned that the technology relieved stress while driving.
- Numerous studies have shown that ACC is well received among adopters primarily because of its perceived convenience and improved safety. However, despite increased usage of ACC while driving, few drivers who own such a system fully understood how the system operated and overestimated the effectiveness of ACC in situations in which it does not work appropriately (Hoedemaeker & Brookhuis, 1998; Jenness, Lerner, Mazor, Osberg, & Tefft, 2008; Llaneras, 2006).
- In a survey conducted by the Automobile Club of Southern California (AAA-FTS, 2008a; Jenness et al., 2008), 370 owners of ACC (out of a total of 1,659 responses from the initial mailed surveys) responded to a questionnaire:
    - Most respondents who have ACC appeared to be satisfied with their systems because the majority of them reported that they would want to purchase ACC again (76%).
    - Although most ACC owners would want to get their system again, many (72%) were not aware of manufacturers' warnings about system limitations.
    - Nearly half of the respondents agreed that using ACC relieves them of stress when driving.
    - Sixteen percent of respondents said that they were "always," "frequently," or "sometimes" confused about whether their ACC system or conventional cruise control system was operating.

Mobility Significance

- No substantive research has been identified that specifically examined the mobility impact of ACC.
    - ACC is conceptually most beneficial for people who primarily drive on highways. An advantage over earlier generation fixed-speed cruise control systems is the ability to function under conditions where traffic speed varies such as conditions of traffic congestion.

Other Benefits

- ACC is estimated to have benefits related to reduced congestion and improved fuel economy due to smoother traffic flow (Marsden, McDonald & Brackstone, 2001).

Technology Penetration

- ACC is primarily available on luxury cars either as standard or optional equipment. However, as more vehicle makes and models begin to feature ACC as either optional or standard features, the price of the system is likely to decline.
- The system has been available in the United States since 2001.

Frequency of Use

- No substantive data was identified on the actual frequency of use of ACC in vehicles so equipped.
- Most systems require the driver to turn on ACC, just as with conventional cruise control. It is not on by default.
- In a recent survey conducted by Eichelberger and McCartt (2012), Volvo drivers were asked whether they use ACC on freeways, expressways, or other high-speed roads.
    - Fifty-one percent reported always using it, while 23% and 5% reported using it sometimes or rarely, respectively.
    - Among those who used ACC, 55% also reported adjusting the gap between the vehicles from the default settings (to either longer (22%) or shorter (33%) following headway time). Whereas 36% never changed the pre-set headway time.

Training and Education

- No formal studies were identified that examined the impact of training/education on the usage of ACC.
- In a survey conducted by the Automobile Club of Southern California (AAA-FTS, 2008a; Jenness et al., 2008), 370 owners of ACC (out of a total of 1,659 responses from the initial mailed surveys) responded to a questionnaire:
    - The most frequently cited method for learning how to use ACC were the vehicle owner's manual and "on-road experience." On-road experience was the only learning method selected by 15.5 percent of respondents.

Behavior Adaptation

- In a test-track study, Rudin-Brown and Parker (2004) assessed whether ACC induces behavioral "adaptation" or over-compensation in drivers in three counterbalanced conditions: No ACC (self-maintained average headway of 2 s), ACC-Short (headway of 1.4 s) and ACC-Long (headway of 2.4 s).
    - Use of ACC resulted in significantly more lane position variability, an effect that was also more pronounced in high sensation-seekers.
    - Driver trust in ACC increased significantly after using the system in this experiment, and these ratings did not change despite a simulated failure of the ACC system during the ACC-Long condition.
- In a simulator study, Xiong et al. (2012) showed that conservative drivers tend to stay farther from the lead vehicle as compared to risky and moderately risky drivers. Risky drivers tended to respond later to critical events and had more ACC warnings.

- In a recent survey conducted by Eichelberger and McCartt (2012), Volvo drivers were asked whether they followed vehicles more or less closely when using ACC. Three percent reported that they followed vehicles more closely, 46 percent followed less closely, and 49 percent reported no change. When asked whether they looked away from the road when using ACC, 4 percent of drivers said they tended to look away from the road more often, 5 percent tended to look away less often, and 90 percent reported no change.
- In a survey conducted by the Automobile Club of Southern California (AAA-FTS, 2008a; Jenness et al., 2008), 370 owners of ACC (out of a total of 1,659 responses from the initial mailed surveys) responded to a questionnaire:
    - Eleven percent of respondents said they usually have their ACC set to the shortest gap (following distance) and 24 percent said that they usually use the longest gap setting.
    - Many ACC owners were not aware of the limitations of their system and overestimate its effectiveness at helping them to avoid collisions. In fact, 72 percent of respondents said that they were not aware of any manufacturer's warnings or limitations about their ACC system.
    - Thirty-eight percent of ACC owners thought that using ACC made them safer drivers than using only conventional cruise control and 7 percent thought that it made them less safe. A majority (54 percent) thought that using ACC made them neither more nor less safe.
    - 12 respondents (3.7%) reported having a collision or "close call" while driving another vehicle equipped with conventional cruise control because they expected the vehicle they were driving to automatically slow down.

Auditory Demand

- Some ACC systems provide an auditory alarm if the driver needs to take action or if the system is disabled. However, no substantive research specifically considering auditory demand associated with ACC use has been identified to date.
    - "The ACC will automatically disengage and send an audio alert of termination at a speed of 40.3 km/h or lower. When the leading vehicle brakes hard and the required deceleration rate exceeds the ACC maximum rate (0.3 g), an audio alert of deceleration limit exceedance is also sent."(Xiong et al., 2012)

Visual Demand

- No substantive research was identified that examined the impact of ACC on visual.

Haptic Demand

- No substantive research was identified that examined the impact of ACC on haptic demand.

Cognitive Demand

- No substantive research was identified that examined the impact of ACC on cognitive demand.

Vehicle Type

- No substantive research was identified that examined the impact of vehicle type on the usability or effectiveness of ACC.

Limitations / Failure Conditions

- ACC only responds to changes in the speed of the forward vehicle. It is not intended to respond to people, animals, stationary obstacles, stopped/parked vehicles on the road, or oncoming and crossing traffic.
- An AAA-FTS assessment (AAA-FTS, 2008b) of results from a survey covering ACC saw benefit potential in the technology, but also raised the concern that many drivers are not aware of the limitations of systems.
    - Misunderstandings identified in the survey included the incorrect assumption that ACC technology would help avoid a collision with a stopped vehicle.
- Some vision based systems can be hindered by rain, fog, and darkness. Also obstruction of the windshield by ice/frost, snow, or dirt can impair sensor function.
- Radar based systems can also be obstructed by snow/ice or dirt/mud which might block the sensor. ACC may not respond well to dirty vehicles that do not reflect enough light or in poor weather conditions.

Differences between Implementations

- Major differences between implementations were not identified during the course of the review (however, this question was not researched in depth).
- Some systems include a lead vehicle graphic in the instrument cluster (or other indicator) to indicate the status of the system. If the indicator is not illuminated, this indicates that the system is not detecting a vehicle in front. This can be useful in indicating to the driver whether the system is functioning properly; if a lead vehicle is present and the indicator is “off”, this provides a cue that the sensor may be damaged or obscured due to dirt or other sources.

References

AAA-FTS. (2008a). Use of Advanced In-Vehicle Technology by Young and Older Early Adopters. Washington, DC. 73.

AAA-FTS. (2008b). Use of Advanced In-Vehicle Technology by Young and Older Early Adopters: Survey Results on Adaptive Cruise Control Systems. In AAA Fundation for Traffic Safety (Ed.), (pp. 2). Washington, DC.

Eichelberger, A. H., & McCartt, A. T. (2012). Volvo Drivers' Experiences with Advanced Crash Avoidance and Related Technologies. Insurance Institute for Highway Safety. Arlington, VA.

Hoedemaeker, M., & Brookhuis, K. (1998). Behavioural adaptation to driving with an adaptive cruise control (ACC). *Transportation Research Part F : Traffic Psychology and Behaviour, 1*, 95-106.

Jenness, J. W., Lerner, N. D., Mazor, S., Osberg, J. S., & Tefft, B. C. (2008). Use of advanced in-vehicle technology by young and older early adopters: Survey results on adaptive cruise control systems. National Highway Traffic Safety Administration. Washington, DC. *DOT HS 810 917*, 105.

Koziol, J., Inman, V., Carter, M., Hitz, J., Najm, W., Chen, S., Lam, A., Penic, M., Jensen, M., Baker, M., Robinson, M., & Goodspeed, C. (1999). Evaluation of the Intelligent Cruise Control System Volume I – Study Results. National Highway Traffic Safety Administration. Washington, D.C. *DOT HS 808 969*, 310.

Llaneras, R. E. (2006). Exploratory study of early adopter, safety-related driving with advanced technologies. for NHTSA by Westat. *DOT HS 809 972*.

NHTSA. (2013). Releases Policy on Automated Vehicle Development, Preliminary Statement of Policy Concerning Automated Vehicles. U.S. Department of Transportation.

Rudin-Brown, C. M., & Parker, H. A. (2004). Behavioural adaptation to adaptive cruise control (ACC): implications for preventive strategies. *Transportation Research Part F : Traffic Psychology and Behaviour, 7*, 59-76.

Xiong, H., Boyle, L. N., Moeckli, J., Dow, B. R., & Brown, T. L. (2012). Use patterns among early adopters of adaptive cruise control. *Human Factors, 54*(5), 722-733.

Other Key References Considered

Braitman, K. A., McCartt, A. T., Zuby, D. S., & Singer, J. (2010). Volvo and Infiniti drivers' experiences with select crash avoidance technologies. *Traffic Inj Prev, 11(3)*, 270-278.

Larsson, A. F. (2012). Driver usage and understanding of adaptive cruise control. *Applied Ergonomics, 43(3)*, 501-506.

Markvollrath, Schleicher, S., & Gelau, C. (2011). The influence of cruise control and adaptive cruise control on driving behaviour--a driving simulator study. *Accident Analysis and Prevention, 43(3*), 1134-1139.

Marsden, G., McDonald, M., & Brackstone, M. (2001). Towards an understanding of adaptive cruise control. *Transportation Research Part C : Emerging Technologies, 9*, 33-51.

Piccinini, G. F., Simoes, A., Rodrigues, C. M., & Leitao, M. (2012). Assessing driver's mental representation of Adaptive Cruise Control (ACC) and its possible effects on behavioural adaptations. *Work, 41 Suppl 1,* 4396-4401.

Seppelt, B. D., & Lee, J. D. (2007). Making adaptive cruise control (ACC) limits visible. *Int. J. Hum.-Comput. Stud., 65(3),* 192-205.

Stanton, N. A., & Young, M. S. (2005). Driver behaviour with adaptive cruise control. *Ergonomics, 48(10),* 1294-1313.

Young, M. S., & Stanton, N. A. (2004). Taking the load off: investigations of how adaptive cruise control affects mental workload. *Ergonomics*, *47(9),* 1014-1035.

Adaptive Headlights

Initial Ratings	**Overall**	**Scenario Specific**
Potential Overall Benefit	☆☆	☆☆☆☆☆
Benefit Currently Documented	★	★★★

What is the technology?

- Adaptive headlights adjust their direction and intensity to provide additional illumination on curves, turns, and hills and to highlight potential hazards.
- Adaptive headlights (adaptive front lighting systems) are lighting devices that adjust the characteristics of the headlight beams to different situations, based upon the steering of the driver. These can apply to either the low beam or high-beam setting.
- Adaptive headlights are a level 0 vehicle automation system (No-Automation) (NHTSA, 2013).

Crash Reduction/Prevention

- An Insurance Institute for Highway Safety (IIHS) study estimated that adaptive headlights have the potential to prevent up to 142,000 crashes per year associated with poor visibility negotiating dark curves (Jermakian, 2011).
    - Estimating the significance of this technology on scenarios related to improving visibility when negotiating curves in darkness or twilight, the study estimated that adaptive headlights have theoretical relevance to 90% of the crashes that occur on curves at night - 91% for nonfatal injury crashes and 88% for fatal crashes.
    - These situations for which there may be potential benefits represent 2% of all crashes, 4% of nonfatal injury crashes, and 8% of fatal crashes.
- The Highway Loss Data Institute (HLDI) looked at adaptive headlights offered by Acura, Mazda, Mercedes and Volvo and found that property damage liability claims fell up to 10% for vehicles with adaptive headlights compared to vehicles without adaptive headlights (IIHS, 2012). Discussions with IIHS (Lund, 2014) indicate that this number would best be translated into a high-end estimate of around a 2.5 to 5% reduction in overall crash events, i.e. there are on the order of two property damage liability claims for a crash event that involves two vehicles. These findings highlight some of the complexity of interpreting insurance data because the IIHS reports also show that:
    - Adaptive headlights do not seem to be associated with a statistically significant reduction in insurance collision claims for Acura, Mercedes and Volvo, the kind of claim that would result from a single-vehicle crash (with the possible exception of Mazda).
    - However, vehicles with adaptive headlights are responsible for fewer crashes with other vehicles, as indicated by a reduction in property damage liability claims and in claims for injuries in other vehicles (IIHS, 2012; Lund, 2013).

    - It's possible that differences between the adaptive headlights and conventional headlights—for example, brightness or beam pattern—may have played a role in reducing crashes with other vehicles.

Consumer Awareness & Trust

- A recent survey of 2,506 Canadians on major available safety technologies (Robertson, Vanlaar, Marcoux, & McAteer, 2012) found that:
    - Only 30.6% of those surveyed were familiar with adaptive headlights (Female: 24.1%, Male: 37.7%).
    - Of those surveyed who were familiar with adaptive headlights, 59.6% agreed that the safety feature offered more protection to passengers in the event of a collision, whereas only 40.5% of those who were not familiar agreed.
- In a study conducted by Sullivan, Flannagan, & Schoettle (2002), participants were not aware of the bending feature of a prototype system of adaptive headlights. Even when asked leading questions by the researcher, participants were not very aware of the bending feature.
- A survey was conducted by the Automobile Club of Southern California with customers (1,117 respondents) who own vehicles that may have high-intensity discharge (HID) headlights or directionally adaptive headlights (Jenness, Lerner, Mazor, Osberg, & Tefft, 2008). The results showed that :
    - Drivers do not necessarily know what type of light source their headlights use. For example, 18 percent of survey respondents did not know whether they had HID headlights and 20 percent did not know whether they had adaptive headlights. Women were more likely than men to say that they didn't know, and older respondents were more likely than younger respondents to say that they didn't know.

Mobility Significance

- Braitman et al. (2010) found that drivers using adaptive headlights reported they were more likely to drive at night.
- In terms of expected impact:
    - Adaptive headlights are expected to be beneficial on moderate- to high-speed roads that are curved and dark. This technology should logically allow drivers with reduced night vision (common in older drivers) to expand the hours in which he/she would feel comfortable driving with the increased visibility provided by adaptive headlights.
    - Adaptive headlights should logically benefit other motorists on the road. For example, when turning around a bend in low-light conditions, standard headlights will temporarily point directly at oncoming traffic, causing glare to oncoming drivers.
        - Unlike standard headlights, adaptive headlights are designed to point more at the road rather than the other driver, thereby reducing the likelihood that

oncoming motorists experience glare from the headlights of others (TIRF, 2013).

    - However, effects of reduced glare are not always consistent across all curve/ turn types)(McLaughlin, Hankey, Green, & Larsen, 2004a).

Other Benefits

- Adaptive headlights have been reported to increase the visibility of pedestrians on unlit curves by 14 percent (Sivak, Flannagan, Traube, Aoki, & Sayer, 1994).
    - Other studies have shown similar results (McLaughlin, Hankey, Green, & Larsen, 2004b). However, it is reported that different implementations might differ in detection distances depending on the curve and turn scenarios (ex. Left vs. right curves and radius of the curves).

Technology Penetration

- Adaptive headlights are primarily available on luxury cars in today's market either as an option or standard. The systems have already begun to appear in modestly priced vehicles (e.g., Mazda 3). They are also available as an aftermarket system. However, as more vehicle makes and models begin to feature adaptive headlights as either optional or standard features, the price of the system is likely to decrease.

Frequency of Use

- No substantive research was identified on the frequency of use of Adaptive Headlights by customers who have that feature on their vehicles.
- Most systems require the driver to turn the light switch to the automatic setting to make this feature available.
- Adaptive headlights automatically disengage when a vehicle is stationary or moving in reverse, so drivers do not need to be concerned about having to turn the feature on and off.
    - That being said, the driver can still turn off the adaptive headlight feature in most systems by moving the lighting switch from AUTO to OFF. There is also typically an adaptive headlight indicator light on the vehicle's dashboard to remind the driver of whether the system is active.
    - For more detailed instructions about turning adaptive headlights on or off, drivers may consult the owner's manual of their vehicles (TIRF, 2013).

Training and Education

- No formal studies were identified that examined the impact of training/education on the usage of Adaptive Headlights.

Behavior Adaptation

- A survey was conducted by the Automobile Club of Southern California with customers (1,117 respondents) who own vehicles that may have high-intensity discharge (HID) headlights or directionally adaptive headlights (Jenness et al., 2008). The results showed that :
    - Nearly a quarter of both older and younger respondents with HID headlights said they are willing to drive faster with their headlights as compared to conventional headlights, and when asked how their driving behavior would change if their HID headlights were replaced with conventional headlights, nearly 18 percent of respondents said they would drive more slowly at night.
- Braitman et al. (2010) found that drivers using adaptive headlights reported they were more likely to drive at night and at higher speeds; whether this translates into any increased risk has not been established.
    - Research on reflector posts, raised pavement markers, and other roadway markings on curves has shown that drivers sometimes increase their speeds when visibility is improved (Zador, Stein, Wright, & Hall, 1987)
- It has been suggested that drivers who put "too much faith" in these systems may be less observant or drive more aggressively (Braitman, McCartt, Zuby & Singer, 2009).

Auditory Demand

- There is no a priori reason to consider auditory demand associated with this technology.

Visual Demand

- No substantive research has been identified that examines the impact of adaptive headlights on visual roadway demand, though presumably the system would increase the amount of visual information while the driver is turning. However, this would be expected to have a beneficial effect, because in nighttime driving visual information is impoverished compared to daytime driving.

Haptic Demand

- There is no a priori reason to consider haptic demand associated with this technology.

Cognitive Demand

- There is no a priori reason to consider specific cognitive demand associated with this technology.

Vehicle Type

- Some systems can swivel the main beams left and right up to 15 degrees, depending on the vehicle's travel path (angle of the curve) and speed (Transport Canada, 2011). This swivel amount provides greater lighting to the road ahead.
- Some systems can automatically switch from high beam to low beam when an approaching vehicle is detected (Transport Canada, 2011).

- There are systems that can shine light 90-degrees in either direction when the vehicle is turning at an intersection. These systems usually use Bi-Xenon or High Intensity Discharge (HID) lights (Transport Canada, 2011).

Limitations / Failure Conditions

- While adaptive headlights can significantly increase a driver's range of visibility, this range still has limits. The system is not designed to alert drivers of nearby obstacles or potential road hazards (Robertson et al., 2012).
- Driving a vehicle equipped with adaptive headlights does not make speeding around corners any safer beyond providing improved illumination; drivers are urged to respect the posted speed limits and to reduce speed appropriately when going around curves (Robertson et al., 2012).

Differences between Implementations

- Major differences between implementations were not identified during the course of the review (however, this question was not researched in depth).

References

Braitman, K. A., McCartt, A. T., Zuby, D. S., & Singer, J. (2010). Volvo and Infiniti drivers' experiences with select crash avoidance technologies. *Traffic Inj Prev, 11*(3), 270-278.

IIHS. (2012). They're working: insurance claims data show which new technologies are preventing crashes. *Status Report, 47*(5), 1-7.

Jenness, J. W., Lerner, N. D., Mazor, S., Osberg, J. S., & Tefft, B. C. (2008). Use of Advanced In-Vehicle Technology By Young and Older Early Adopters Survey Results on Headlamp Systems. National Highway Traffic Safety Administration. Washington, DC. *DOT HS 810 902*, 107.

Jermakian, J. S. (2011). Crash avoidance potential of four passenger vehicle technologies. *Accident Analysis and Prevention, 43*(3), 732-740.

Lund, A. (2013). *Drivers and Driver Assistance Systems: How well do they match*? Presentation at the 7th International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design, Bolton Landing, NY, USA, June 18, 2013.

Lund, A. (2014). Personal communication expanding upon interpretation of data presented in Lund (2013) and other IIHS reports; April 3, 2014.

McLaughlin, S., Hankey, J., Green, C. A., & Larsen, M. (2004a). Discomfort Glare Ratings of Swiveling HID Headlamps. SAE International. Warrendale, PA. *2004-01-2257*.

McLaughlin, S., Hankey, J., Green, C. A., & Larsen, M. (2004b). Target Detection Distances and Driver Performance with Swiveling HID Headlamps. *SAE TECHNICAL Paper Series, 2004-01-2258*.

NHTSA. (2013). Releases Policy on Automated Vehicle Development, Preliminary Statement of Policy Concerning Automated Vehicles. U.S. Department of Transportation.

Robertson, R. D., Vanlaar, W. G. M., Marcoux, K., & McAteer, H. J. (2012). Vehicle safety features: knowledge, perceptions, and driving habits. Traffic Injury Research Foundation. 77.

Sivak, M., Flannagan, M. J., Traube, E. C., Aoki, M., & Sayer, J. (1994). Evaluation of an active headlight system. University of Michigan Transportation Research Institute. Ann Arbor, MI. *Report no. UMTRI-94-17*.

Sullivan, J. M., Flannagan, M. J., & Schoettle, B. (2002). The Appearance of Bending Light From Other Vehicles. The University of Michigan Transportation Research Institute. *UMTRI-2002-2*, 28.

TIRF. (2013). http://brainonboard.ca/. Retrieved 2013-09-16

Transport Canada. (2011, 2011-06-20). http://www.tc.gc.ca/eng/roadsafety/safevehicles-1186.htm. Retrieved 2013-09-16

Zador, P., Stein, H. S., Wright, P., & Hall, J. (1987). Effects of chevrons, post-mounted delineators, and raised pavement markers on driver behaviour at roadway curves. *Transportation Research Record, 11114,* 1-10.

Other Key References Considered

http://www.iihs.org/iihs/topics/t/crash-avoidance-technologies/qanda#crash-avoidance-technologies

http://www.tc.gc.ca/eng/roadsafety/safevehicles-1186.htm

Back-Up / Rear-View Cameras

Initial Ratings	**Overall**	**Scenario Specific**
Potential Overall Benefit	☆	☆☆☆
Benefit Currently Documented	★	★★

Note: The relatively modest ranking for Back-Up cameras in terms of Overall Safety Benefit is a result of the relatively low number of backup event related injuries and fatalities relative to the total number of driving related injuries and fatalities. It is also recognized that there is a particularly high emotional cost associated with this type of event. NHTSA (2014) estimates that 31% of all backup event fatalities involve children under 5 years of age and another 26% are adults 70 years and older; these events often involve family members or other close associations. Societal pressure to do something about such events led to NHTSA recently issuing a final rule on March 31st of this year mandating rear visibility technology in all new vehicles under 10,000 pounds by May 2018.

What is the technology?

- Back-Up Cameras allow the driver to view a video image of the area behind the rear bumper and see small objects that are not ordinarily visible with mirrors or looking over the shoulder.
  - The video is displayed either on a screen in the instrument panel or in a corner of the inside rearview mirror, and is automatically activated when the transmission is shifted into reverse.
- Back-Up Cameras are a level 0 vehicle automation system (No-Automation) (NHTSA, 2013).

Crash Reduction/Prevention

- Based on 2007 NiTS (Not-in-Traffic Surveillance) data, NHTSA estimated that 221 deaths and 14,000 injuries occur annually in non-traffic backover crashes. In addition, an average of 71 backover fatalities and 4,000 backover injuries are reported each year on public roadways (NHTSA, 2008), resulting in a combined total of 292 deaths and 18,000 injuries.
  - 103 of these estimated 292 annual deaths involved children younger than 5, and 76 deaths involved people 70 and older. About 2,000 of the 18,000 injuries that occur every year from backover crashes involve children younger than 5, and 3,000 involve people 70 and older.
  - These figures likely underestimate the frequency of backover crashes, as their lower severity makes them less likely to be reported (i.e. little property damage, minor injuries).
- NHTSA (2010) released proposed rules in the Federal Register that estimated that annual fatalities occurring from backing crashes could be reduced from 207 to 112 (46% reduction) if all vehicles were equipped with rearview video technology. The estimate for reduction in annual injuries was from 15446 to 8374 (46%).
- In experimental settings, Mazzae (Mazzae, 2008, 2010, 2013) found that the use of back-up cameras reduced crashes in an unexpected collision trial by approximately 30% across all three studies.

- In a controlled, surprise collision event after 4 weeks of Backup Camera use, only 5 out of 13 drivers tested received a rear collision warning from the installed system, suggesting that these systems may not be universally reliable in all types of backing scenarios (Mazzae, 2008).

- An experimental IIHS research study (Kidd, Hagoski, Tucker, & Chiang, 2014) using an SUV and conducted with volunteers suggests that rearview camera systems would aid in preventing more backover crashes into pedestrians in a vehicle's rear blind zone than rear proximity (parking) warning sensors. Perhaps surprisingly, the research found that cameras alone worked better than the combination of both rearview camera and backup warning sensors.

- Data from the Highway Loss Data Institute (2012) considering an initial assessment of Mercedes vehicles with and without backup cameras showed small and mixed findings across insurance claims and damages. The report concluded that the data showed no significant effect on any insurance coverage; however, this was considered a relatively weak analysis for injury effects involving pedestrians and it was stated that additional analyses were underway. An initial analysis considering Mazda vehicles (HLDI, 2011) found that, contrary to expectations, there was an increase in collision frequency claims (3.1%), severity, and overall losses ($18), but a non-significant reduction in property damage / liability claims. Most relevant from a safety perspective, there was a reduction in the frequency of high severity bodily injury claims of 22.2%.

Consumer Awareness & Trust

- A survey conducted by the Automobile Club of Southern California on Sensor-Based Backing Systems and Rear-View Video Cameras showed that (Jenness, Lerner, Mazor, Osberg, & Tefft, 2007):
    - Most respondents (96%) found their camera to be easy or very easy to use when backing out of a driveway.
    - Approximately 36 percent of respondents agreed or strongly agreed with the statement that "the rear-view camera does not show the entire area behind my vehicle that I need to see when backing, in other words there is a blind spot."
    - Older respondents (aged 65 and older) were more likely than younger respondents (younger than 65) to say that they would want to get the system again.
    - Only 39 percent of rear-view camera owners reported that they were aware of "any warnings or limitations" about their system. The percentage of respondents who said that they were aware of warnings or limitations varied significantly by vehicle manufacturer. Also, a higher percentage of younger backing aid owners (26%) as compared to older owners (18%) were aware of system limitations.

- A survey of nearly 300 drivers who recently purchased vehicles (model years 2000-2004) with parking aids (proximity sensors or back-up cameras) for backing was conducted to assess their opinions and use of the devices (NHTSA, 2006).
    - Approximately 25% of the cars sampled were equipped with a rearview camera system.

- Eighty percent of drivers thought the parking aid (back-up cameras + proximity sensors) would lower their likelihood of being involved in a backing-related crash, but a few drivers (11 percent) believed that the system might increase the likelihood.
- Sixty-seven percent of owners believed that their parking aid system would provide warnings at any backing speed. However, most systems only operate at speeds less than 6 mph. Thus, while this survey indicates that many drivers like the systems and find them helpful, some drivers had beliefs that might lead to decreased safety in some circumstances.

Mobility Significance

- No substantive research was identified in the course of this review that specifically examined the impact of back-up cameras on mobility significance.
    - Conceptually, back-up cameras might be expected to be beneficial for drivers with physical limitations (such as inability to turn their heads sufficiently to look over the shoulder) that make inspections towards the rear of the car difficult. This generally becomes more of an issue as drivers age.
    - As was noted earlier, a AAA survey (Jenness, Lerner, Mazor, Osberg, & Tefft, 2007) found that Older respondents (aged 65 and older) who had experience with a vehicle with a back-up camera were more likely than younger respondents (younger than 65) to say that they would want this technology in their next car. This could be interpreted as indirect support the perceived utility of the technology and possible mobility relevance in older drivers.

Other Benefits

- None Identified.

Technology Penetration

- Back-Up Cameras are available for numerous new vehicles on the market today as an add-on option or a standard feature. They are also available as an aftermarket system. However, as more vehicle makes and models begin to feature Back-Up Cameras as either optional or standard features, the price of the system is likely to decline (NHTSA, 2006).
    - It is expected that back-up cameras will reach 42.4 million units by 2020, up from just 11.4 million units in 2012, with a compound annual growth rate CAGR) of 19.6 percent (IHS iSuppli's, 2013).

Frequency of Use

- Most systems automatically turn on the Back-Up Cameras when the car is put into reverse.
- No substantive data has been identified to date on the frequency of use for Back-Up Cameras.

Training and Education

- A review of relevant research indicated that factors such as prior experience with rearview cameras, expectations regarding the likelihood of an obstacle during backing, and the timing of glances to the camera images influence the use and subsequent benefits of these systems (Llaneras, Neurauter, & Green, 2011).
- A survey conducted by the Automobile Club of Southern California on Sensor-Based Backing Systems and Rear-View Video Cameras (Jenness et al., 2007) found that older drivers were more likely to have learned how to operate their systems from the owner's manual. (A higher percentage of younger respondents learned to use their systems from on-road experience and practice.)

Behavior Adaptation

- As reported in (Jenness et al., 2007), a survey conducted by the Automobile Club of Southern California on sensor-based backing systems and rear-view video cameras found that approximately 17 percent of rear-view camera owners admitted backing without checking their mirrors or turning to look out the rear window within the last two weeks. Younger system owners were more likely have done this than were older system owners.
    - It has been hypothesized by some that drivers who have systems that pair back-up cameras with a proximity warning alarms may be less careful than drivers who do not have these system when driving in reverse (e.g., not looking out for pedestrians or small children; using the mirrors less; making fewer shoulder checks; and driving in reverse with more speed than without the system) because they assume that the alarm will system will serve this function for them. (It should not be assumed that these same individuals would have appropriately completed these types of safety checks if their vehicles were not equipped with these technologies – further research on this question seems warranted.)
- As noted above, an experimental IIHS research study (Kidd, Hagoski, Tucker, & Chiang, 2014) found that back-up cameras alone worked better than the combination of both rearview camera and backup warning sensors.
    - IIHS researchers have suggested that the back-up sensors may have given drivers a false sense of security such that they paid less attention to the camera display. It was noted that slightly fewer drivers who had both systems operational looked at the camera display at least once than participants who had only the camera display. In addition, drivers with the combined system spent a smaller proportion of time looking at the display; however, these differences were not statistically significant and should be interpreted cautiously.

Auditory Demand

- There is no a priori reason to consider Back-Up Camera technology alone to be a source of added auditory demand.
    - Only systems that are paired with proximity sensor alarms, which typically issue auditory alarms when objects are detected, would impose an auditory demand on the driver.

Visual Demand

- Conceptually, a Back-up camera system might be considered as a technology that reduces some aspects of visual demand by possibly improving visibility directly behind the vehicle. At the same time, visual load might be considered to have been increased in some ways when there is an expectation that the driver should both make use of the camera display as well as continuing to visually inspect mirrors, looking over the shoulder to check behind the vehicle, etc. No objective data has been identified to date as part of this review that specifically addresses this issue.

Haptic Demand

- There is no a priori reason to consider Back-Up Camera technology alone to be a source of added haptic demand, since no haptic alarm is produced by systems on the market today.

Cognitive Demand

- No research data has been identified to date in the course of this review that specifically addresses the question of cognitive demand.

Vehicle Type

- More SUVs and pickup trucks than cars are involved in backover crashes (NHTSA, 2008).
    - SUVs and pickup trucks typically have bigger blind zones than cars because they sit higher off the ground, making it more difficult for drivers to see children and smaller objects near the rear of the vehicle (Mazzae & Barickman, 2009; Mazzae & Garrott, 2008).
- An analysis of driveway backovers involving children in Utah from 1998-2003 found that children were 53% more likely to be injured by a pickup truck than a car and 2.4 times more likely to be injured by a minivan, relative to the number of registered vehicles of each type (Pinkney et al., 2006).
    - Future studies should control not just for the number of registered vehicles for each type, but also for the fact that minivans are more likely to be owned by families with children.
- Data from the Highway Loss Data Institute (2011) show that Back-Up Cameras are infrequently installed in non-hybrid vehicles (19%), but are very common in hybrids (83%).

Limitations / Failure Conditions

- Hurwitz et al. (2010) evaluated the use of rearview cameras with sensor systems. Thirty-five drivers completed 16 parking trials. Only seven drivers looked at the camera image before backing. Of those who didn't, 46% looked at it after the sensor issued an audible warning.
    - This is a classic issue with some forms of support technology – while a technology such as a back-up camera offers the potential to aid the driver, the driver cannot benefit from this type of technology unless they make use of it.
- In a controlled, surprise collision event after 4 weeks of Backup Camera use, only 5 out of 13 drivers tested received a rear collision warning from the installed system, suggesting that these systems may not be universally reliable in all types of backing scenarios (Mazzae, 2008).
- Conditions such as rain, darkness, glare and/or dirt on the camera lens could make visibility more difficult than without such conditions.

Differences between Implementations

- Some systems offer interactive track lines on the video screen that turn along with the steering wheel to help direct your path.
    - Fixed guidelines: Show the actual path of the vehicle while reversing in a straight line, which can be helpful when backing into a parking space or aligning the vehicle with another object behind the vehicle.
    - Centerline: Helps align the center of the vehicle with an object (e.g. a trailer).
    - Active guidelines: Shows the intended path of the vehicle when reversing.
- Some systems are equipped with sensors such as radar or ultrasonic systems to warn the driver of objects behind the vehicle or of vehicles approaching from the sides.Some systems will even automatically apply the brakes to keep the vehicle from backing into or over an object.
- Mazzae (2010) found that a rear collision warning system was as effective in reducing crashes as a full camera system, and that the camera systems were most effective when embedded in the rearview mirror.
- In many vehicles, blind zones could be reduced through better vehicle designs that increase the directly viewable area (Hammond & Wade, 2005), that would be anticipated to increase safety even without a backup camera – although the issue remains that a driver had to actually look to benefit.
- Different systems may have different fields of view and induce differing amounts of optical distortion in the rearview image.
- Image quality may vary between systems, depending on the exact type of camera used (i.e. black and white vs. color camera, color depth, contrast and resolution).

References

Hammond, C., & Wade, M. G. (2005). Forward looking blindspots: A report of A-Pillar induced field-of-view obstruction and driver performance in a simulated rural environment. *Advances in Transportation Studies: An International Journal, Section B5,* 69-81.

HLDI. (2011). Mazda collision avoidance features: initial results. Highway Loss Data Institute. *28.13.*

HLDI. (2012). Mercedes-Benz collision avoidance features: initial results Highway Loss Data Institute. *29.7.*

Hurwitz, D., Pradhan, A. K., Fisher, D., Knodler, M., Muttart, J., Menon, R., & Meissner, U. (2010). Backing collisions: a study of drivers' eye and backing behavior using combined rear-view camera and sensor systems. *Injury Prevention, 16*, 79-84.

IHS iSuppli's. (2013, May 16). Cameras in Motor Vehicles to Grow More Than Fivefold by 2020. Retrieved December 6, 2013, from http://www.isuppli.com/Automotive-Infotainment-and-Telematics/MarketWatch/Pages/Cameras-in-Motor-Vehicles-to-Grow-More-Than-Five-fold-by-2020.aspx

Jenness, J. W., Lerner, N. D., Mazor, S. D., Osberg, J. S., & Tefft, B. C. (2007). Use of Advanced In-Vehicle Technology By Young and Older Early Adopters: Results on Sensor-Based Backing Systems and Rear-View Video Cameras. National Highway Transportation Safety Administration. Washington, DC. *DOT-HS-810-828rev*, 178.

Kidd, D.G., Hagoski, B.K., Tucker, T.G., & Chiang, D.P. (2014). Effects of a rearview camera, parking sensor system, and the technologies combined on preventing a collision with an unexpected stationary or moving object. Insurance Institute for Highway Safety, Arlington, VA.

Llaneras, R. E. (2006). Exploratory study of early adopter, safety-related driving with advanced technologies. for NHTSA by Westat. *DOT HS 809 972*.

Llaneras, R. E., Neurauter, M. L., & Green, C. A. (2011). *Factors moderating the effectiveness of rear vision systems: what performance-shaping factors contribute to unexpected in-path obstacles when backing?* Paper presented at the SAE 2011 World Congress & Exhibition, Warrendale, PA.

Mazzae, E. N. (2008). On-Road Study of Drivers' Use of Rearview Video Systems (ORSDURVS). National Highway Traffic Safety Administration. Washington, D.C. *No. DOT HS 811 024*, 140.

Mazzae, E. N. (2010). Drivers' Use of Rearview Video and Sensor-Based Backing Aid Systems in a Non-Laboratory Setting. National Highway Traffic Safety Administration. Washington, D.C. *No. NHTSA-2010-0162*, 16.

Mazzae, E. N. (2013). Rearview Video System Use by Drivers of a Sedan in an Unexpected Obstacle Event National Highway Traffic Safety Administration. Washington, D.C. *No. NHTSA-2010-0162*, 39.

Mazzae, E. N., & Barickman, F. (2009). Direct rear visibility of passenger cars: Laser-based measurement development and findings for late model vehicles. National Highway Traffic Safety Administration. Washington, DC. *DOT HS-811-174*.

Mazzae, E. N., & Garrott, W. R. (2008). Light vehicle rear visibility assessment. National Highway Traffic Safety Administration. Washington, DC. *DOT HS-810-909.*

NHTSA. (2006). Vehicle Backover Avoidance Technology Study. U.S. Department of Transportation. Washington, DC. 65.

NHTSA. (2008). Fatalities and injuries in motor vehicle backing crashes. U.S. Department of Transportation. Washington, DC. *DOT HS 811 144*.

NHTSA. (2013). Releases Policy on Automated Vehicle Development, Preliminary Statement of Policy Concerning Automated Vehicles. U.S. Department of Transportation.

NHTSA. (2014). Federal Motor Vehicle Safety Standards; Rear Visibility; Final Rule. U.S. Department of Transportation. Federal Register, 79(66),

Pinkney, K., Smith, A., Mann, N., Mower, G., Davis, A., & Dean, J. (2006). Risk of pediatric back-over injuries in residential driveways by vehicle type. *Pediatric Emergency Care, 6*, 402-407.

Other Key Sources Reviewed

http://www.tc.gc.ca/eng/roadsafety/safevehicles-backing-aids-1069.htm

https://www.media.volvocars.com/global/enhanced/en-gb/Media/Preview.aspx?mediaid=49569

http://www.cnn.com/2013/09/25/us/new-cars-rearview-video-systems/index.html?hpt=hp_t2

Anthikkat, A. P., Page, A., & Barker, R. (2013). Low-speed vehicle run over fatalities in Australian children aged 0-5 years. Journal of Pediatrics and Child Health, 49(5), 388-393.

Ayres, T., Li, L., Trachtmanc, D., & Young, D. (2005). Passenger-side rear-view mirrors: driver behavior and safety. International Journal of Industrial Ergonomics, 35, 157-162.

Consumer Report. (2012). Best and worst rear blind zones. Retrieved August 4, 2012, from http://www.consumerreports.org/cro/2012/03/the-danger-of-blind-zones/index.htm

Griffin, B., Watt, K., Wallis, B., Shields, L., & Kimble, R. (2011). Paediatric low speed vehicle run-over fatalities in Queensland. Injury Prevention, 17(Supplement), i10-13.

IIHS. (2011). Comment to the National Highway Traffic Safety Administration concerning proposed amendments to rearview mirrors safety standard (Vol. NHTSA-2010-0162). Arlington, VA.

Kidd, D. G. & Brethwaite, A. (2013). Visibility of children behind 2010-13 model year passenger vehicles using glances, mirrors, and backup cameras and parking sensors. Insurance Institute for Highway Safety, Arlington, VA.

Neeman, T., Wylie, J., Attewell, R., Glase, K., & Wallace, A. (2002). Driveway Deaths: Fatalities of Young Children in Australia as a Result of Low-Speed Motor Vehicle Impacts (Vol. Road Safety Report no. CR208). Canberra, Australia: Australian Transport Safety Bureau.

Nhan, C., Rothman, L., Slater, M., & Howard, A. (2009). Back-over collisions in child pedestrians from the Canadian hospitals injury reporting and prevention program. Traffic and Injury Prevention, 10(4), 350-353.

NHTSA. (2010). Federal Motor Vehicle Safety Standard, Rearview mirrors; Federal Motor Vehicle Safety Standard, Low-Speed Vehicles Phase-In Reporting Requirements; proposed rule (Vol. 75(234)).

Roberts, I., Norton, R., & Jackson, R. (1995). Driveway-related child pedestrian injuries: a case-control study. Pediatrics in Review, 95, 405-408.

U.S. House of Representatives. (2007). Cameron Gulbransen Kids Transportation Safety Act of 2007. Washington, DC: US Congress.

Forward Collision Warning (FCW)

Initial Ratings	**Overall**	**Scenario Specific**
Potential Overall Benefit	☆☆☆	☆☆☆☆☆
Benefit Currently Documented	★	★★

What is the technology?

- Forward Collision Warning (FCW) is designed to warn drivers of possible hazards in front of their vehicles so that they can take braking or steering actions to avoid crashes or reduce the damage from unavoidable crashes. (Forward Collision Mitigation (FCM) systems that actively brake or steer the vehicle are considered separately.)
    - The system monitors the relative speed and following distance from the forward vehicle, or the distance to an unmoving object if it is estimated to be in the forward path of the vehicle. When the combination of speed and distances (i.e., the time headway or the time to collision) becomes critical, a signal (audible, haptic, visual, or some combination) is presented to alert the driver.
- The area in front of the car is monitored by a sensor (e.g., RADAR, LIDAR, and/or camera).
- Forward Collision Warning is a level 0 vehicle automation system (No-Automation)(NHTSA, 2013) .

Crash Reduction/Prevention

- Early simulator based research suggested that FCW can redirect the driver's attention to the road and improve reaction time (Lee, McGehee, Brown, & Reyes, 2002).
- In a mathematical simulation drawing on data from the National Automotive Sampling System Crashworthiness Data System (NASS CDS), Kusano and Gabler (2012) estimated a 3.2% rear-end crash prevention benefit for FCW.
    - Extending this analysis to possible injury reduction, the combined FCW technology alone was estimated to potentially prevent 29% of Abbreviated Injury Scale (AIS) injuries of category 2 or above.
    - Modeling using the German In-Depth Accident Study (GIDAS) database and effectiveness estimates based on reactions to a Bosch system by researches affiliated with the manufacturer (Georgi et al. 2009) estimated that FCW feature would translate into a safety benefit of a 38% reduction in rear-end crashes.
- In a simulation study (Yasuda et al. 2011) estimated that FCW could reduce rear-end crashes at a relative velocity of 20 km/hr by 30%.
- FCW has since been examined at least one field operational test (Najm, Stearns, Howarth, Koopmann, & Hitz, 2006). A prototype FCW system without automatic braking was field tested by 66 drivers for four weeks each. Based on the number of near-crash scenarios identified, the system was projected to reduce rear-end collision rates by 10 percent.

- Property damage liability claim rates are lower than average for vehicles equipped with FCW (Lund, 2013). The same analyses found that vehicle models that also equipped with autonomous braking (forward collision mitigation) (i.e. Acura and Mercedes) are associated with even lower rates than the same vehicle models with only FCW.
    - IIHS has stated in follow-on discussions (Lund, 2014) that the available property damage loss data shows a reduction in claims for vehicles equipped with FCW in the range of 5 to 7%. IIHS estimates that this translates into a 10 to 15% reduction in rear crashes.

Consumer Awareness & Trust

- In a recent survey (Robertson, Vanlaar, Marcoux, & McAteer, 2012), 2,506 Canadians were polled on major available safety technologies (832 over the phone and 1,674 online).
    - The results showed that only 23.6% of the respondents were familiar with FCW (Female: 15.1%, Male: 32.6%).
    - 60.0% of drivers who were familiar with FCW said that FCW would be easy to use (Female: 54.1%, Male: 66.2%).
    - 46.5% of respondents considered FCW to be useful.
    - 66.3% of those surveyed mentioned that they would be willing to use FCW if it is already included in the car.
- In a large field study (Sayer et al., 2011), drivers rated the usefulness and satisfaction with FCW lowest among the subsystems evaluated.
    - Overall, drivers rated them neutral with regard to satisfaction, but recognized that they had some utility.
    - The brake pulse accompanying FCWs was the single system attribute that drivers disliked most.

Mobility Significance

- No substantive research was identified during the course of this review that explicitly considered the impact of FCW on mobility.
    - It might be hypothesized that safety technologies such as FCW that are intended to alert drivers of potentially dangerous situations may increase the confidence of individuals who are otherwise concerned about their ability to continue to drive safely due to aging or other factors; however, this needs to be objectively evaluated.

Other Benefits

- None Identified

Technology Penetration

- FCW is primarily available on luxury cars in the current U.S. vehicle fleet. However, as more vehicles begin to feature FCW as either optional or standard features, the price of the system is likely to decline and its market penetration increased.

Frequency of Use

- No substantive data has been identified to date on the frequency with which drivers with cars equipped with FCW drive with or without the technology active.
- Most FCW systems are on by default, and require drivers to deactivate them if they do not want to use them.

Training and Education

- No formal studies were identified that examined the impact of training/education on the usage of FCW.

Behavior Adaptation

- In a field study (Sayer et al., 2011) evaluating the impact of an integrated crash warning system found that drivers were slightly more likely to maintain shorter headways; more time was spent at time headways of one second or less with the integrated system in the treatment condition (24%) than in the baseline condition (21%).
  - Some have suggested that drivers who put too much faith in these systems may be less observant or drive more aggressively; however, additional research is indicated to determine whether this may or may not be a substantive behavioral pattern and whether safety net safety benefits are negatively impacted as a result.

Auditory Demand

- Many FCW systems issue an auditory alarm. However, no substantive research specifically considering auditory demand associated with FCW use has been identified to date.

Visual Demand

- FCW systems may include a visual indicator and/or alarm. However, no substantive research specifically considering visual demand associated with FCW use has been identified to date.

Haptic Demand

- Several of field trials have used systems that include haptic stimuli; however, no substantive research specifically considering haptic demand associated with FCW use has been identified to date.

Cognitive Demand

- No substantive research has been identified to date that specifically has considered the cognitive demand associated with FCW systems.

Vehicle Type

- No relevant data on vehicle type has been identified in to date; however, this question has not been a significant focus of the literature review at this point.

Limitations / Failure Conditions

- While a Forward Collision Warning system may provide assistance in directing a driver's attention to a potential collision event, drivers need to be aware that FCW technology still requires an appropriate response by the driver to avoid or mitigate the severity of a potential crash and that the technology is not designed to or able to warn of all potential crash situations. Drivers need to maintain an appropriate level of attention to the driving environment at all times even when driving with an FCW system active.

- Camera-based systems are less effective at night than radar-based systems and can be "blinded" by direct sun light (e.g., early sunrise and late sunset).
- The effectiveness of both radar and camera based systems can be compromised by snow/ice build-up in front of the sensors.

Differences between Implementations

- Some FCW implementations automatically prepare the brake system for rapid braking (prime the brakes) when an alarm is active. The system does not automatically activate the brakes but, if the brake pedal is pressed, full force braking is applied even if the brake pedal is lightly pressed.
    - The FCW's brake support can only help reduce the speed at which a collision occurs if the driver applies the vehicle's brakes. The brake pedal still must be pressed, as in a typical braking situation.
    - FCW is thus a warning only, and is to be distinguished from Forward Collision Mitigation, where the system may actually apply the brakes.

References

Georgi, A., Zimmermann, M., Lich, T., Blank, L., Kickler, N., & Marchthaler, R. (2009). New approach of accident benefit analysis for rear end collision avoidance and mitigation systems. Proceedings of the 21st International Technical Conference on the Enhanced Safety of Vehicles, Stuttgart, Germany.

Kusano, K. D. & Gabler, H.C. (2012). Safety Benefits of Forward Collision Warning, Brake Assist, and Autonomous Braking Systems in Rear-End Collisions. *IEEE Transactions on Intelligent Transportation Systems, 13(4),* 1546-1555.

Lee, J. D., McGehee, D. V., Brown, T. L., & Reyes, M. L. (2002). Collision warning timing, driver distraction, and driver response to imminent rear-end collisions in a high-fidelity driving simulator. *Human Factors, 44*(2), 314-334.

Lund, A. (2013). *Drivers and Driver Assistance Systems: How well do they match*? Presentation at the 7th International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design, Bolton Landing, NY, USA, June 18, 2013.

Lund, A. (2014). Personal communication expanding upon interpretation of data presented in Lund (2013) and other IIHS reports; April 3, 2014.

Najm, W. G., Stearns, M. D., Howarth, H., Koopmann, J., & Hitz, J. (2006). Evaluation of an automotive rear-end collision avoidance system. National Highway Traffic Safety Administration. Washington, DC. (DOT HS-810-569).

NHTSA. (2013). Releases Policy on Automated Vehicle Development, Preliminary Statement of Policy Concerning Automated Vehicles. U.S. Department of Transportation.

Robertson, R. D., Vanlaar, W. G. M., Marcoux, K., & McAteer, H. J. (2012). Vehicle safety features: knowledge, perceptions, and driving habits. Traffic Injury Research Foundation. 77.

Sayer, J. R., Bogard, S. E., Buonarosa, M. L., LeBlance, D. J., Funkhouser, D. S., Bao, S., Blankespoor, A. D., & Winkler, C. B. (2011). Integrated vehicle-based safety systems light-vehicle field operational test key findings report. National Highway Traffic Safety Administration. Washington, DC.

Yasuda, H., A. Kozato, et al. (2011). A forward collision warning (FCW) performance evaluation. Proceedings of the 22nd International Technical Conference on the Enhanced Safety of Vehicles. Washington, DC.

Other Key Sources Reviewed

http://www.tc.gc.ca/eng/roadsafety/safevehicles-1183.htm

Bao, S., LeBlanc, D. J., Sayer, J. R., & Flannagan, C. (2012). Heavy-truck drivers' following behavior with intervention of an integrated, in-vehicle crash warning system: a field evaluation. *Human Factors, 54(5),* 687-697.

Ben-Yaacov, A., Maltz, M., & Shinar, D. (2002). Effects of an in-vehicle collision avoidance warning system on short- and long-term driving performance. *Human Factors, 44(2),* 335-342.

Brackstone, M., & McDonald, M. (2007). Driver headway: How close is too close on a motorway? *Ergonomics, 50(8),* 1183-1195.

Wege, C., Will, S., & Victor, T. (2013). Eye movement and brake reactions to real world brake-capacity forward collision warnings-A naturalistic driving study. *Accident Analysis and Prevention, 58,* 259-270.

Forward Collision Mitigation (FCM) / Collision Imminent Braking / Autobrake

Initial Ratings	**Overall**	**Scenario Specific**
Potential Overall Benefit	☆☆☆	☆☆☆☆☆
Benefit Currently Documented	★★	★★★

What is the technology?

- This technology class may be identified by a range of alternate names. The term used in this document is Forward Collision Mitigation (FCM), but other frequently used terms include collision imminent braking, autobrake, and autonomous braking.
- FCM systems detect the distances and closing speeds of objects in the path of the vehicle and automatically decelerate or stop the vehicle if the driver does not respond to the alarm provided by the system.
    - Crashes that are potentially preventable may still occur due to late braking and/or braking without sufficient force. Many drivers are not used to dealing with safety-critical braking situations and do not apply enough braking force to avoid a crash. FCM is designed to reduce the number and severity of these types of collisions.
- Implementations of this class of technology may include Forward Collision Warning (FCW) and Brake Assist (BA) technology that pre-primes the brake system.
- Forward Collision Mitigation is a level 1 vehicle automation system (Function-Specific Automation) (NHTSA, 2013).

Crash Reduction/Prevention

- An Insurance Institute for Highway Safety (IIHS) study has estimated that FCW/FCM could prevent up to 1.2 million crashes in the United States each year, including 66,000 serious and moderate injury crashes and 879 fatal crashes (Jermakian, 2011).
    - In terms of events, this would translate into potentially preventing or mitigating approximately 20% of the total number of police-reported traffic crashes.
    - In terms of fatalities, this would represent a little over 2% of total deaths.
- An Australian report employing a crash reconstruction technique (Anderson et al. 2012) that considered 104 crashes involving rear-end and other FCM relevant scenarios and estimated a 20-40% reduction in fatal crashes and 30-50% of injury crashes if 100% penetration of the technology was assumed.
- In a mathematical simulation drawing on data from the National Automotive Sampling System Crashworthiness Data System (NASS CDS), Kusano and Gabler (2012) estimated that combined FCW and pre-crash brake assist technologies could have prevented 7.7% of the rear-end collisions modeled.
    - Extending this analysis to possible injury reduction, the combined FCM technologies were estimated to potentially prevent 50% of Abbreviated Injury Scale (AIS) injuries of category 2 or above.

- Another modeling approach combining data from NASS CDS and test track and simulation data (Van Auken, Zellner et al. 2011) and developing estimates for a next generation FCM system predicted that installation of the technology throughout the light-vehicle fleet might decrease crash events by 9.3% and fatalities by 3.7%.
    - Estimates for rear-end crashes were reduction of events by 28.1% and fatalities by 35.1%.
- A statistical modeling / simulation evaluation based on the German In-Depth Accident Study (GIDAS) database and supplemented by speed characterization and driver behavior data from NASS CDS and Volvo's crash database (Coelingh et al. 2007) estimated that, assuming ideal conditions such as100% penetration, dry road conditions, etc., a 50% reduction in rear-end crashes might result.
    - A subsequent study with the same lead author (Coelingh et al. 2010) considering additional combinations of technologies as part of the FCM system and somewhat different modeling and data reported an estimated reduction in fatalities for rear-end crashes of 30%.
    - Another study using the GIDAS database and estimates based on driver reactions to a Bosch system (Georgi et al. 2009) reported potential benefits of a 55% reduction in rear-end crashes for all drivers and up to a 72% reduction when considering an implementation with automated emergency braking for drivers who failed to respond to an emergent braking situation.
- In a simulator based study (Yasuda et al. 2011), in which vehicles were proceeding at a relative velocity of 20 km/h, rear-end crash reduction estimates of 48% were reported for FCW plus brake assist and an extremely high estimation of a 90% reduction in rear-end crashes under these conditions with a system that include FCW, brake assist, and automated emergency braking.
- Property damage liability claim rates are lower on average for vehicles equipped with forward collision mitigation (Eichelberger & McCartt, 2012; IIHS, 2012; Lund, 2013). Moreover, the models equipped with autonomous braking (i.e. Acura and Mercedes) are more effective than similar vehicles equipped only with forward collision warning.
    - An insurance claims based study based on comparable Volvo models with and without a FCM system (Isaksson-Hellman & Lindman, 2012) and reported a 23% reduction in rear-end crashes for the equipped vehicles.
- The Insurance Institute for Highway Safety (IIHS) conducted a series of five test runs at speeds of 12 and 25 mph on the track at the Vehicle Research Center (IIHS, 2013). In each test, an engineer drove the vehicle toward a stationary target designed to simulate the back of a car. Sensors in the test vehicle monitored its lane position, speed, time to collision, braking and other data.
    - The highest-scoring cars and SUVs have autobrake and substantially reduce speeds in both the 12 and 25 mph tests. Most of these systems prevent the 12 mph collision. However, not all the cars behave the same and this is why cars have been categorized by the IIHS as a three-tier rating system of superior, advanced

and basic to reflect that even a basic forward collision warning system can provide significant benefits.

Consumer Awareness & Trust

- In a recent survey (Robertson, Vanlaar, Marcoux, & McAteer, 2012), 2,506 Canadians were polled on major available safety technologies (832 over the phone and 1,674 online). The following results were for Forward Collision Warning (FCW) but may have some relevance for FCM:
  - The results showed that only 23.6% of the respondents were familiar with FCW (Female: 15.1%, Male: 32.6%).
  - 60.0% of drivers who were familiar with FCW said that FCW would be easy to use (Female: 54.1%, Male: 66.2%).
  - 46.5% of respondents considered FCW to be useful.
  - 66.3% of those surveyed mentioned that they would be willing to use FCW if it is already included in the car.
- A 2012 IIHS survey (Eichelberger & McCartt, 2012) of owners of Volvo vehicles with crash avoidance technologies including FCM (City Safety) found that, despite some annoyance: the majority of drivers left the systems turned on most of the time; felt the systems made them safer drivers; and would want them in their next vehicle.
- False braking events can adversely affect customer confidence and acceptance in FCM.
  - A false 0.6 g braking event is more likely to provoke a negative customer reaction compared to a simple alarm (loud beeping noise and flashing red warning lights).

Mobility Significance

- No substantive research was identified during the course of this review that explicitly considered the impact of FCW on mobility.
  - It might be hypothesized that safety technologies such as FCM that are intended to alert and, if necessary, intervene for drivers in potentially dangerous situations may increase the confidence of individuals who are otherwise concerned about their ability to continue to drive safely due to aging or other factors; however, this needs to be objectively evaluated.

Other Benefits

- None Identified

Technology Penetration

- FCM is primarily present in luxury vehicles today. However, as more vehicles begin to feature FCM as either optional or standard features, the price of the system is likely to decrease.

Frequency of Use

- No substantive data has been identified to date on the frequency with which drivers with cars equipped with FCM drive with or without the technology active.
- Most FCM systems are on by default, and require drivers to deactivate them if they do not want to use them.

Training and Education

- No substantive publicly available data or research was identified that examines the impact of training/education on the usage of Forward Collision Mitigation.

Behavior Adaptation

- In a field study (Sayer et al., 2011) evaluating the impact of an integrated crash warning system found that drivers were slightly more likely to maintain shorter headways; more time was spent at time headways of one second or less with the integrated system in the treatment condition (24%) than in the baseline condition (21%).
    - Some have suggested that drivers who put too much faith in these systems may be less observant or drive more aggressively; however, additional research is indicated to determine whether this may or may not be a substantive behavioral pattern and whether safety net safety benefits are negatively impacted as a result.

Auditory Demand

- Many FCM systems issue an auditory alarm. However, no substantive research specifically considering auditory demand associated with FCW use has been identified to date.

Visual Demand

- FCM systems may include a visual indicator and/or alarm. However, no substantive research specifically considering visual demand associated with FCW use has been identified to date.

Haptic Demand

- Several of field trials of FCW / FCM have used systems that include haptic stimuli; however, no substantive research specifically considering haptic demand associated with FCW use has been identified to date.

Cognitive Demand

- No substantive research has been identified to date that specifically has considered the cognitive demand associated with FCM systems.

Vehicle Type

- No relevant data on vehicle type has been identified in to date; however, this question has not been a significant focus of the literature review at this point.

Limitations / Failure Conditions

- While currently available test track data (IIHS, 2013) indicate that some FCM systems are likely to offer substantial safety benefits under certain conditions, drivers need to be aware that FCM technology will not prevent all forms of crashes and that they need to maintain an appropriate level of attention to the driving environment at all times.
- Weather and environmental conditions can influence the system. Camera-based systems are less effective at night than radar-based systems. Also, camera-based systems can be "blinded" by direct sun light (e.g., early sunrise and late sunset). Both radar and camera systems can be obscured by snow/ice build-up in front of the sensors.

Differences between Implementations

- A recent experiment conducted by the IIHS (IIHS, 2013)showed that manufacturers have implemented their FCM systems differently.
    - Some FCM systems can slow down or completely stop the car to avoid some front-to-rear crashes if the driver fails to brake or steer out of the way in response to a warning; others were much less effective under the conditions tested.
- Some systems, depending on the conditions, are also designed to detect parked vehicles, stationary vehicles, and other roadside objects such trees, guard rails, sign posts, etc.
- Some systems are designed to detect pedestrians and large animals in daylight conditions.

References

Anderson, R. W. G., Doecke, S. D., Mackenzie, J. R., Ponte, G., Paine, D., & Paine, M (2012). Potential benefits of forward collision avoidance technology. Australia, Centre for Automotice Safety Research (Report CASR106).

Coelingh, E., A. Eidehall, Bengtsson, M. (2010). Collision Warning with Full Auto Brake and Pedestrian Detection -- a practical example of Automatic Emergency Braking. 13th International IEEE Annual Conference on Intelligent ransportation Systems, Madeira Island, Portugal.

Coelingh, E., L. Jakobsson, L., Lind, H., & Lindman, M. (2007). Collision Warning with Auto Brake - A Real-Life Safety Perspective. 20th International Technical Conference on the Enhanced Safety of Vehicles (ESV), Lyon, France.

Eichelberger, A. H., & McCartt, A. T. (2012). Volvo Drivers' Experiences with Advanced Crash Avoidance and Related Technologies. Insurance Institute for Highway Safety. Arlington, VA.

Georgi, A., Zimmermann, M., Lich, T., Blank, L., Kickler, N., & Marchthaler, R. (2009). New approach of accident benefit analysis for rear end collision avoidance and mitigation systems. Proceedings of the 21st International Technical Conference on the Enhanced Safety of Vehicles, Stuttgart, Germany.

Isaksson-Hellman, I. & Lindman, M. (2012). The Effect of a Low-Speed Automatic Brake System Estimated from Real Life Data. 56th AAAM Annual Conference Annals of Advances in Automotive Medicine.

IIHS. (2012). They're working: insurance claims data show which new technologies are preventing crashes. *Status Report, 47*(5), 1-7.

IIHS. (2013, September 27, 2013). First crash avoidance ratings under new test program: 7 midsize vehicles earn top marks.   Retrieved October 28, 2013, from http://www.iihs.org/iihs/sr/statusreport/article/48/7/1

Kusano, K. D. & Gabler, H.C. (2012). Safety Benefits of Forward Collision Warning, Brake Assist, and Autonomous Braking Systems in Rear-End Collisions. *IEEE Transactions on Intelligent Transportation Systems, 13(4),* 1546-1555.

Jermakian, J. S. (2011). Crash avoidance potential of four passenger vehicle technologies. *Accident Analysis and Prevention, 43*(3), 732-740.

Lee, J. D., McGehee, D. V., Brown, T. L., & Reyes, M. L. (2002). Collision warning timing, driver distraction, and driver response to imminent rear-end collisions in a high-fidelity driving simulator. *Human Factors, 44*(2), 314-334.

Lund, A. (2013). *Drivers and Driver Assistance Systems: How well do they match*? Presentation at the 7th International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design, Bolton Landing, NY, USA, June 18, 2013.

NHTSA. (2013). Releases Policy on Automated Vehicle Development, Preliminary Statement of Policy Concerning Automated Vehicles. U.S. Department of Transportation.

Robertson, R. D., Vanlaar, W. G. M., Marcoux, K., & McAteer, H. J. (2012). Vehicle safety features: knowledge, perceptions, and driving habits. Traffic Injury Research Foundation. 77.

Sayer, J. R., Bogard, S. E., Buonarosa, M. L., LeBlance, D. J., Funkhouser, D. S., Bao, S., Blankespoor, A. D., & Winkler, C. B. (2011). Integrated vehicle-based safety systems light-vehicle field operational test key findings report. National Highway Traffic Safety Administration. Washington, DC.

Van Auken, R. M., Zellner, J.W., Chiang, D.P., Kelly, J., Silberling, J.Y., Dai, R., Broen, P.C., Kirsch, A.M., & Sugimoto, Y. (2011). Advanced Crash Avoidance Technologies (ACAT) Program - Final Report of the Honda-DRI Team, Volume I: Executive Summary and Technical Report. Washington, DC, US DOT NHTSA: 285.

Yasuda, H., A. Kozato, et al. (2011). A forward collision warning (FCW) performance evaluation. Proceedings of the 22nd International Technical Conference on the Enhanced Safety of Vehicles. Washington, DC.

Other Key Sources Reviewed

http://www.bosch-automotivetechnology.com/en/com/component_com/SF_PC_DA_Predictive-Emergency-Braking-System_SF_PC_Driver-Assistance-Systems_2567.html?compId=2568

Balint, A., Fagerlind, H., & Kullgren, A. (2013). A test-based method for the assessment of pre-crash warning and braking systems. *Accident Analysis and Prevention, 59,* 192-199.

Chien, S., Li, L., & Chen, Y. P. (2012). A new braking and warning scoring system for vehicle forward collision imminent braking systems. Paper presented at the 15th International IEEE Conference on Intelligent Transportation Systems, Anchorage, Alaska, USA.

Isaksson-Hellman, I., & Lindman, M. (2012). The effect of a low-speed automatic brake system estimated from real life data. *Ann Adv Automot Med, 56,* 231-240.

Matsui, Y., Han, Y., & Mizuno, K. (2011). Performance of collision damage mitigation braking systems and their effects on human injury in the event of car-to-pedestrian accidents. *Stapp Car Crash J, 55,* 461-478.

Page, Y., Foret-Bruno, J., & Cuny, S. (2005). Are expected and observed effectiveness of emergency brake assist in preventing road injury accidents consistent? Paper presented at the 19th International Technical Conference on Enhanced Safety of Vehicles, Washington, D.C.

Rosen, E., Kallhammer, J. E., Eriksson, D., Nentwich, M., Fredriksson, R., & Smith, K. (2010). Pedestrian injury mitigation by autonomous braking. *Accident Analysis and Prevention, 42(6),* 1949-1957.

Terry, H. R., Charlton, S. G., & Perrone, J. A. (2008). The role of looming and attention capture in drivers' braking responses. *Accident Analysis and Prevention, 40(4),* 1375-1382.

Lane Departure Warning (LDW)

Initial Ratings	**Overall**	**Scenario Specific**
Potential Overall Benefit	☆☆☆	☆☆☆
Benefit Currently Documented	★	★

What Is the Technology?

- Lane Departure Warning (LDW) systems provide alerts whenever a driver appears to (i.e. a turn signal is not activated) drift too close to the edges of the lane or partially cross the lane marking (i.e. when a turn signal is not activated). They are primarily designed to reduce high speed accidents on highways and freeways. (Note: this review does not consider "lane keeping assist" / lane departure prevention technologies that adjust steering to actively assist in keeping the vehicle within lane boundaries.)
- LDW systems typically use visual, audible, and/or vibratory warnings.
- LDW only systems should be differentiated from lane keeping assistance systems that automatically adjust steering to keep the vehicle in the lane. Lane keeping assistance systems may provide a warning before engaging.
- LDW is a level 0 vehicle automation system (No-Automation)(NHTSA, 2013)

Crash Reduction/Prevention

- Crash records were extracted from the 2004-2008 files of the National Automotive Sampling System General Estimates System (NASS GES) and the Fatality Analysis Reporting System (FARS) (Jermakian, 2011). Crash descriptions were reviewed to determine whether the information or action provided by each technology potentially could have prevented or mitigated the crash. This technology potentially could prevent/mitigate up to 179,000 crashes per year (3% of all crashes), including 37,000 nonfatal serious and moderate injury crashes and 7,529 fatal crashes.
  - Based on the published 25–30% effectiveness estimates for rumble strips, a more realistic estimate of crashes that may be prevented by LDW systems would be 45,000–54,000 per year (Jermakian, 2011).
  - A statistical estimation study using crash data from the United Kingdom (Robinson et al, 2011) and estimates of technology effectiveness reported a benefit potential in the range of 7-29% for fatality reduction and 13-34% for serious injuries.
  - A study using crash data from the Australian state of New South Wales (Anderson et al. 2011) developed benefit estimates of fatal crash reductions in scenario specific conditions in the range of 11-13% and reductions of 1-9% for injuries.
- A field operational test conducted in 2004 and 2005 (Wilson, Stearns, Koopmann, & Yang, 2007) estimated that LDW systems could reduce road-departure crashes by between 9,400 and 74,800 annually if all passenger vehicles were equipped with these systems and they worked as intended.

    - During the field test, LDW systems were turned off 45% of the time because of weather and other factors. At that rate, the estimated reduction in crashes would drop to between 5,200 and 41,200.
- A project using statistical modeling and simulation of a Volvo pre-production LDW system (Gordon et al, 2010) produced an estimated benefit of a 47% reduction in lane-departure crashes, assuming ideal conditions.
    - A likely more realistic evaluation considering variable lane markings, non-ideal weather condtions, etc., produced an estimated target crash reduction of 33%. Taking further factors into consideration, a final estimated crash rate reduction was placed in the range of 13% to 31%.
- A simulation modeling study that drew on an analysis of a database of serious road-departure crashes (Kusano & Gabler, 2012) found a wide range of estimated benefit depending on when warnings were provided. Estimated crash reductions were 3-5% for warnings delivered at time of lane-crossing and 19-34% for warnings delivered one second prior.
    - Another modeling based simulation study (Tanaka et al. 2012), but based on Japanese crash data, also found potential benefit to be dependent upon the timing of warnings. In a model where the warning was delivered one second after a lane crossing resulted in an estimated 5% reduction in scenario specific crashes. Maximum benefit was seen in the modeling for warnings provided 1 second prior to a lane-crossing, which showed a 25% reduction.
- In a field study (Sayer et al., 2011), an integrated system (LDW plus curve-speed warning (CSW), lane-change/ merge (LCM), and FCW) had a statistically significant effect on the frequency of lane departures, decreasing the rate from 14.6 departures per 100 miles during baseline driving (two weeks without active LDW) to 7.6 departures per 100 miles during treatment (period in which the system was operational). (See also report by Nodine et al. (2011).)
    - Keeping in mind that the system included features beyond basic LCW, Nodine et al. (2011) estimated that the combined technologies had a target crash type reduction potential in the range of 6% - 29%.
- In contrast with estimates from FOT and simulation modeling studies, a recent analysis of insurance claim data (Lund, 2013) suggests that LDW alone (without active lane keeping assistance) may actually increase these types of crashes.
    - Only Volvo LDW systems were associated with decreases in claim frequency, and that is likely due to pairing of LDW with forward collision warning and forward collision mitigation technologies (autonomous braking systems).

Consumer Awareness & Trust

- In a recent survey conducted by the Traffic Injury Research Foundation (TIRF), 2,506 Canadians completed a poll on major available safety technologies (832 over the phone and 1,674 online) (Robertson, Vanlaar, Marcoux, & McAteer, 2012).
    - Most respondents were not familiar with LDW systems. Only 21.6% of respondents were aware of this system (15.9 for women, 28.3% for men).
    - Of those who were familiar with LDW, 47.9% agreed that the safety feature in new vehicles help protect drivers in the event of a collision. For those who were not familiar with LDW, 37.3% agreed. (These endorsements are somewhat concerning given that LDW alone is a warning system and should not logically provide any actual protection in a crash event except to the extent to which it might reduce the severity of a crash if an evasive maneuver undertaken in response to the warning reduced the severity of a crash event.)

Mobility Significance

- No substantive research has been identified to date that examined the mobility significance of lane departure warning systems.
    - It may be hypothesized that appropriate implementations of this technology might increase driver confidence and enhance comfort while driving for all drivers, including those with disabilities; however, this needs to objectively evaluated.

Other Benefits

- None Identified

Technology Penetration

- No substantive research has been identified to date examining the technology penetration of LDW systems.

Frequency of Use

- As noted earlier, in a field test (Wilson, Stearns, Koopmann, & Yang, 2007) LDW systems were found to be turned off 45% of the time because of weather and other factors.
- Of Volvo and Infinity drivers who own vehicles equipped with LDW and consented to a telephone interview (Braitman, McCartt, Zuby, & Singer, 2010):
    - 69% reported using the system all of the time
    - 93% reported using their LDW systems at least occasionally
    - It should be kept in mind that the extent to which reported use and actual use correspond is unknown in such self-report data
- A subsequent interview study of Volvo owners (Eichelberger & McCartt, 2012) found a lower self-reported frequency of use, with 59% reporting that they used the system all of the time. Other findings included:

- One-quarter (25%) of owners of LDW systems mentioned that the warnings were annoying, and 9% said they were distracting. For all the other technologies considered in the interview (adaptive cruise control, distance alert, collision warning with full auto brake, and driver alert control), fewer than 5% mentioned that those technologies were annoying or distracting.
- Sixty-six percent (66%) of owners who were annoyed said they had turned off the system, compared with 19% of those who were not annoyed or did not know whether they were annoyed.

Training and Education

- No substantive research has been identified to date that examined the impact of training/education on the usage of LDW.

Behavior Adaptation

- During a field test (LeBlanc et al., 2006), drivers using LDW systems improved their lane-keeping behavior, traveling near or beyond the lane edge less frequently, and increased their use of turn signals.
- Self-report data on usage must be interpreted cautiously if observational data is not available for comparison purposes. See the survey data above under "Frequency of Use" as well as the following:
    - 67% of survey respondents said that LDW improved their general lane-keeping behavior, and 60% said they increased use of turn signals (Braitman et al., 2010).
    - IIHS conducted a survey of Volvo owners (Eichelberger & McCartt, 2012) who drove with LDW turned on. Fifty-five percent reported no change in their use of the turn signal when the system was turned on, and 44% of drivers said they used their turn signal more often with it. Sixty-one percent reported no change in how often they drifted from their lane, and 35% said they drifted from their lane less often.
- In a field study (Sayer et al., 2011) with an integrated system that included LDW, drivers were less likely to make unsignaled lane changes in the treatment condition than during baseline driving and had a lower number and reduced duration of lane excursions. See also report by Nodine et al. (2011).

Auditory Demand

- Owners who had heard a LDW auditory alert were surveyed (Eichelberger & McCartt, 2012). Among these drivers, 96% agreed that the warning sound was useful, 33% agreed it was annoying, 7% agreed it was too loud, and 1% agreed the sound was too quiet.

Visual Demand

- LDW systems usually do not use a visual warning, instead relying on auditory alarms and/or haptic (vibration feedback) in the steering wheel or seat. When visual alerts are involved, typical warnings may include a flashing symbol on the dashboard display or heads-up display

Haptic Demand

- Some LDW systems vibrate the steering wheel or the driver's seat (on the side corresponding to the lane departure.) to simulate a rumble strip-like sensation.

Cognitive Demand

- No substantive research has been identified that examined the cognitive demands of LDW systems.

Vehicle Type

- In some systems, LDW has two settings: a less sensitive setting that warns the driver when a tire crosses a lane marking, and a more sensitive setting that warns the driver before the tire crosses the lane marking. The less sensitive setting is the default setting.
- Some systems require lane markings on only one side of the vehicle, whereas other systems require lane markings on both sides (Eichelberger & McCartt, 2012).
- Some LDW systems incorporate road edge detection which attempts to identify unmarked transitions between the pavement and a soft shoulder.

Limitations/Failure Conditions

- In a field test of a prototype road departure warning system, the system was available 76% of the time on freeways compared with only 36% on non-freeways (Wilson et al., 2007).
- In a survey of Volvo owners, 77% reported that the LDW system had never failed to warn them when they believed they were at risk of drifting out of their lane. However, 17% reported that it had (Eichelberger & McCartt, 2012). That is, the system failed to warn them when it should have (a "false negative"). The most frequently reported situations in which this happened included missing or unclear lane markings (60%), inclement weather (17%), driving at slow speeds (7%), and driving in the dark (7%).
- An area which may be worthy of more research is the issue of driver annoyance with false alarms, which may be particularly an issue with LDW systems. See Tijerina et al. (2010) for a consideration of an adaptive LDW design which was intended to reduce frustration by reducing false alarms. The trade-off with such a system is that the number of times the system should warn the driver but fails to then increases.
- Lane marking quality and environmental conditions can affect LDW performance to the extent that they impact the system's ability to identify traffic lanes (see citation above).
    - Many systems rely upon having good, visible road markings and may not be able to detect an unmarked road edge.
    - Lane departure warning systems rely on the ability of the sensors to register lane markings, which may be problematic on roads that are not well marked or are covered with snow.
    - Many current LDW systems have reduced lane detection rates in low light or inclement weather compared to normal light and weather conditions (Sayer et al., 2010).

    - Sensors such as cameras may be influenced by environmental factors such as lighting or precipitation. Wilson et al. (2007) found that an LDW system was available 56% of the time during dry, daytime conditions, but only 4% of the time during wet, nighttime conditions.
- Many crash avoidance technologies rely on the driver to take action (IIHS, 2012). The effectiveness of these systems depends on whether drivers accept the technologies, understand the information from the system, and respond appropriately. This is especially true for warning systems, since a valid warning is useless if it is ignored. It is generally assumed that if drivers experience too many false alarms, they may find the systems to be annoying, overwhelming, or unhelpful and may disable them. Additionally, interpreting warnings from multiple systems may be confusing or even distracting for some drivers. Objective data on these assumptions is limited or lacking all together.

Differences between Implementations

- Different LDW systems may activate at different speeds (Eichelberger & McCartt, 2012), depending on the manufacturer (between 40 mph to 120 mph vs. over 70 km/h).
    - They may also de-activate at different speeds.
- Some LDW systems are combined with input from a Blind Spot Radar System to provide an imminent warning if there is a vehicle in the space you are approaching when executing a lane change.

References

Anderson, R. W. G., Hutchinson, T. P., Linke, B.J., & Ponte, G. (2011). Analysis of crash data to estimate the benefits of emerging vehicle technology (CASR094). Adelaide, Australia, Centre for Automotive Safety Research. The University of Adelaide.

Braitman, K. A., McCartt, A. T., Zuby, D. S., & Singer, J. (2010). Volvo and Infiniti drivers' experiences with select crash avoidance technologies. *Traffic Injury Prevention, 11*(3), 270-278.

Eichelberger, A. H., & McCartt, A. T. (2012). Volvo Drivers' Experiences with Advanced Crash Avoidance and Related Technologies. Insurance Institute for Highway Safety. Arlington, VA.

Gordon, T., Sardar, H., Blower, D., Ljung Aust, M., Bareket, Z., Barnes, M., et al. (2010). Advanced Crash Avoidance Technologies (ACAT) Program – Final Report of the Volvo-Ford-UMTRI Project: Safety Impact Methodology for Lane Departure Warning – Method Development and Estimation of Benefits. Washington, DC: 218.

IIHS. (2012). They're working: insurance claims data show which new technologies are preventing crashes. *Status Report, 47*(5), 1-7.

Jermakian, J. S. (2011). Crash avoidance potential of four passenger vehicle technologies. *Accident Analysis and Prevention, 43*(3), 732-740.

Kusano, K. & Gabler, H. (2012). Model of collision avoidance with lane departure warning in real-world departure collisions with fixed roadside objects. 15th International IEEE Conference on Intelligent Transportations Systems, Anchorage, AK.

LeBlanc, D., Sayer, J., Winkler, C., Ervin, R., Bogard, S., Devonshire, J., Mefford, M., Hagan, M. T., Bareket, Z., Goodsell, R., & Gordon, T. (2006). Road Departure Crash Warning System Field Operational Test: Methodology and Results. University of Michigan Transportation Research Institute. Ann Arbor, MI. (UMTRI-2006-9-1).

Lund, A. (2013). *Drivers and Driver Assistance Systems: How well do they match*? Presentation at the 7th International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design, Bolton Landing, NY, USA, June 18, 2013.

NHTSA. (2013). Releases Policy on Automated Vehicle Development, Preliminary Statement of Policy Concerning Automated Vehicles. U.S. Department of Transportation.

Nodine, E. E., Lam, A.H., Najm, W.G., & Ference, J.J. (2011). Safety impact of an integrated crash warning system based on field test data. 22nd International Technical Conference on the Enhanced Safety of Vehicles. Washington, DC.

Robinson, B., Hulshof, W., Cookson, R., Cuerden, R., Hutchins, R., & Delmonte, (2011). Cost benefit evaluation of advanced primary safety systems. Transport Research Laboratory, Report No. PPR586.

Sayer, J. R., Bogard, S. E., Buonarosa, M. L., LeBlance, D. J., Funkhouser, D. S., Bao, S., Blankespoor, A. D., & Winkler, C. B. (2011). Integrated vehicle-based safety systems light-vehicle field operational test key findings report. National Highway Traffic Safety Administration. Washington, DC. *DOT HS 811 416*.

Sayer, J. R., Bogard, S. E., Funkhouser, D., LeBlanc, D. J., Bao, S., Blankespoor, A. D., Buonarosa, M. L., & Winkler, C. B. (2010). Integrated vehicle-based safety systems heavy-truck field operational test key findings report. National Highway Traffic Safety Administration. Washington, DC. *DOT HS-811-362*.

Tanaka, S., Mochida, T., Aga, M., & Tajima, J. (2012). Benefit estimation of a lane departure warning system using ASSTREET. *SAE International Journal of Passenger Cars – Electronic and Electrical Systems, 5(1),* 133-145.

Tijerina, L., Blommer, M., Curry, R., Greenberg, J., Kochhar, D., Simonds, C., & Watson, D. (2010). Effects of adaptive lane departure warning system on driver response to a surprise event. *Transportation Research Record, 2185(2185),* 1-7.

Wilson, B. H., Stearns, M. D., Koopmann, J., & Yang, C. Y. (2007). Evaluation of a road-departure crash warning system. National Highway Traffic Safety Administration. Washington, DC. (DOT-HS-810-854).

Other Key Sources Reviewed

http://www.tc.gc.ca/eng/roadsafety/safevehicles-1184.htm

Hetrick, S. (1997). Examination of driver lane change behavior and the potential effectiveness of warning onset rules for lane change or “side” crash avoidance systems. (M.Sc.), Virginia Polytechnic Institute and State University, Blacksburg, VA.

Navarro, J., Mars, F., Forzy, J. F., El-Jaafari, M., & Hoc, J. M. (2010). Objective and subjective evaluation of motor priming and warning systems applied to lateral control assistance. *Accident Analysis and Prevention, 42(3),* 904-912.

Peng, Y., Boyle, L. N., & Hallmark, S. L. (2013). Driver's lane keeping ability with eyes off road: Insights from a naturalistic study. *Accident Analysis and Prevention, 50*, 628-634.

Robertson, R. D., Vanlaar, W. G. M., Marcoux, K., & McAteer, H. J. (2012). Vehicle safety features: Knowledge, perceptions, and driving habits (pp. 77): Traffic Injury Research Foundation.

Tapia-Espinoza, R., & Torres-Torriti, M. (2013). Robust lane sensing and departure warning under shadows and occlusions. Sensors (Basel), 13(3), 3270-3298.

Appendix D – Educational Support Materials

The RFP that initiated this project did not specifically call for the development of educational support materials for the consumer on the selected technologies. The MIT AgeLab project leads elected to gather information and descriptions on each of the technologies that might prove useful in developing such support material as desired at a future date. There is no presumption on our part that the representative content provided need be presented in the same format as currently used in the sheets.

The educational support material sheets include the following content sections:

- *What Is It?* – A short one to two paragraph description of what a technology is and the conditions under which it might be relevant. This is intended as a very brief, high level orientation to the technology.

The sections listed below represent a next level down description and elaboration of information on the technology, but again presented at a consumer oriented level.

- *Why Would I Use This Technology?* – This section generally expands somewhat on what the technology is, why it is relevant, and sometimes includes additional information on how and/or why it works.
- *What Do Drivers Think?* – This section generally highlights information on consumer satisfaction with the technology.
- *How Well Does It Work?* – This section addresses objective data on the expected potential and/or observed safety benefits of the technology. It is based on selected data drawn from the *Crash Reduction/Prevention* section of the technology review sheet.
- *Who Benefits Most?* – This section highlights relevant information on the type of drivers, driving conditions, or type of vehicles that may benefit the most from availability of this technology.
- *In What Situations Doesn't It Work*? - Similar to the *Limitations / Failure Conditions* section of the technology review sheet, but presented at a consumer level.
- *Mobility Significance* – Similar to the *Mobility Significance* section of the technology review sheet, but presented at a consumer level.
- *Not All Systems Are Alike* – Similar to the *Differences between Implementations* section of the technology review sheet, but presented at a consumer level.
- *Different Names, Same Idea* – A listing of alternate names that different manufacturers may use to describe their implementation(s) of a particular type of safety technology.

As noted in the main body of the report, these supplemental educational support material sheets are seen as starting points that would benefit from additional review and refinement to ensure fully accurate translation of technical details into consumer oriented language. Citations from which most statements are made are provided in the "Technology Review Sheets" provided in Appendix C and reference to them will likely be useful in reviewing. This material is presented here to support idea generation and to serve as a reference base in the further development of consumer oriented educational support materials.

Electronic Stability Control (ESC)

What Is It?

Electronic Stability Control systems help you maintain or regain control of your vehicle in difficult driving situations, such as during unexpected turns or while negotiating icy roads. These systems work by automatically applying the brakes, reducing the engine power, and/or adjusting the vehicle's tire suspension to prevent "loss of control" events. "Loss of control" accidents are extremely common, accounting for up to 40% of fatal car crashes. Studies have shown that Electronic Stability Control systems are highly effective in helping a driver maintain control of the vehicle, and substantially reduce the risks associated with loss of control, preventing many accidents.

Electronic Stability Control continuously monitors tire movement and steering wheel activity to sense a loss of traction or slippage. In such situations, Electronic Stability Control systems can reduce engine power, apply brakes independently to each wheel, and correct tire suspension much faster than the driver could. These systems are particularly helpful in managing unexpected events, or driving on wet or icy roadways. They are also especially helpful to drivers of large vehicles, such as SUVs.

Just the Facts

Why Would I Use This Technology?

Loss of control crashes, in which the driver cannot adequately control his/her vehicle, are extremely common, accounting for up to 40% of fatal accidents annually. ESC systems detect situations where the vehicle may not be responding to the driver's control, and automatically adjust the brake, throttle and tire suspension to compensate. The primary benefit of this technology is that it **increases the driver's ability to control the vehicle and substantially reduces accident risk.**

What Do Drivers Think?

About 50% of consumers do not know about ESC or whether it is implemented in their vehicle. Those who are familiar with ESC generally have a high opinion of it. In a survey of ESC aware drivers, 89% said that they felt the technology made them safer.

How Well Does It Work?

ESC systems are highly effective, and could help prevent up to 800,000 single-vehicle crashes per year (48%). Fatal crashes could be reduced 40-56%. An experiment that simulated common loss of control situations found that accidents were greatly reduced in the ESC-enabled situations.

Who Benefits Most?

ESC systems are highly beneficial overall, especially for drivers who own larger vehicles such as SUVs. In addition, drivers who commonly find themselves on icy or wet roadways would also derive the most benefit from these systems.

In What Situations Doesn't It Work?

ESC systems are designed to detect situations in which the driver may be under- or over-steering (as in a sharp, sudden turn), or in cases where the driver is not adequately controlling the vehicle (as on icy roads). The system would not affect cases such as unintentionally drifting out of the lane. It does not have much effect on rear-end collisions. Anecdotal reports by some drivers suggests that they may drive more aggressively under some conditions due to their confidence in the system's ability to recover control, which may lower the overall gain that might be obtained otherwise.

Mobility Significance

ESC systems increase driving comfort in adverse driving conditions, reducing driver stress.

Not All Systems Are Alike

ESC is designed to be kept on at all times. Some versions have an "off switch" that can be used in situations where ESC might control the vehicle too often (as when stuck in snow or mud). The type of switch and how long it stays off will vary between car manufacturers.

Different Names, Same Idea

Electronic Stability Control systems can be found under a number of different names, including:

- Vehicle Stability Assist
- Vehicle Dynamic Control
- Electronic Stability Program
- Dynamic Stability Control
- StabiliTrak
- AdvanceTrac

Adaptive (Automatic) Cruise Control (ACC)

What Is It?

Adaptive (Automatic) Cruise Control senses where the forward vehicle is relative to your own vehicle, and slows down and speeds up your vehicle to maintain a consistent headway time. Unlike traditional cruise control, which can only be set to a single speed, Adaptive (Automatic) Cruise Control can adapt when other vehicles change their speed. Adaptive (Automatic) Cruise Control may help prevent around 13,000 crashes per year.

Adaptive (Automatic) Cruise Control is most beneficial to drivers who often drive on highways. It may also improve the driver's ability to navigate traffic and improve fuel economy. It is important to keep in mind that these systems are designed to respond to other moving vehicles, and do not detect objects that are very small or stationary. Camera-based systems can be affected by the time of day and weather conditions, whereas radar-based systems can be obstructed by ice or snow.

<u>Just the Facts</u>

Why Would I Use This Technology?

Adaptive (Automatic) Cruise Control adjusts the vehicle's speed in response to other vehicles' changes in speed. The primary benefit of this technology is that it **helps the driver manage his speed and maintain a safe headway time to other cars.**

What Do Drivers Think?

Drivers who already own an ACC system have a very high opinion of it, with 76-93% of survey respondents reporting that they would buy the system again. Nearly half of respondents said that the system helps relieve stress. About a third of respondents said that the system made them a safer driver.

How Well Does It Work?

ACC systems may help prevent 13,000 crashes per year. Surveys suggest (see above) that most drivers are very satisfied with these systems.

Who Benefits Most?

ACC is primarily available in luxury and higher-end vehicles, though this is expected to change in the coming years. Drivers who often drive on highways would derive the most benefit from this technology.

In What Situations Doesn't It Work?

ACC systems are not designed to respond to stationary or particularly small objects. Surveys have shown that relatively few drivers are aware of these types of limitations, and may overestimate the system's protective benefit. Some drivers also have difficulty telling when ACC is active, as opposed to standard cruise control.

Mobility Significance

Although ACC’s mobility benefits are relatively minor, the system would help smooth a driver’s control of the vehicle, and allow him/her to focus on other aspects of driving.

Not All Systems Are Alike

ACC can be found under several configurations based on the time between the vehicle.

Different Names, Same Idea

Adaptive (Automatic) Cruise Control can be found under a number of different names, including:

- Autonomous cruise control
- Intelligent cruise control

Adaptive Headlights

What Is It?

Adaptive headlights adjust their direction and intensity in response to the driver's steering to provide additional light on curves, turns, hills, or to highlight potential hazards. Poor visibility is a common cause of crashes, and this technology may help prevent up to 142,000 crashes per year (about 90% of all crashes caused by visibility problems).

Adaptive headlights are most effective on moderate to high-speed roads in dark conditions, particularly when going around curves. Adaptive headlights also may be beneficial to persons other than the driver. They point more toward the road, reducing glare for other drivers, and may increase the visibility of pedestrians.

<u>Just the Facts</u>

Why Would I Use This Technology?

Adaptive headlights adjust their direction and intensity to provide additional illumination on curves, turns, and hills and to highlight potential hazards. The primary benefit of this technology is that it **increases the visible range of the forward roadway, particularly when navigating curves and turns.**

What Do Drivers Think?

Studies show that few drivers are aware of adaptive headlamp technology (31%), among those who are, 60% agree that it is a beneficial safety feature.

How Well Does It Work?

The Insurance Institute for Highway Safety estimates that adaptive headlights are potentially relevant to 90% of crashes that occur on curves at night; however, the extent to which having adaptive headlights translates into actual crash reductions is still being actively studied. It has been found that vehicles equipped with adaptive headlights are involved in fewer crashes with other vehicles than vehicles that are not.

Who Benefits Most?

Adaptive headlights would be most beneficial to drivers who often drive at night. Since the technology is passive (automatic), it will be of some use to almost all drivers.

In What Situations Doesn't It Work?

Though these systems can improve visibility, they do not warn the driver about potential obstacles. The lights do not make it safer to speed. This is especially important to keep in mind, as about a quarter of drivers who use adaptive headlights have reported that they are willing to drive faster with this technology turned on.

Mobility Significance

Adaptive headlights enhance the driver's field of view in dark or other poor visibility conditions. They extend the hours during which a person can comfortably drive.

Not All Systems Are Alike

Future adaptive headlamp systems may adjust the beam pattern to prevent glare for the oncoming motorist, thereby allowing your vehicle to operate in high beam mode more frequently. Federal requirements (FMVSS) do not presently permit such systems.

Different Names, Same Idea

Adaptive headlights can be found under a number of different names, including:

- Advanced Forward Lighting System
- Adaptive Headlamp

Additional terms that might be used to identify this technology include "cornering headlights", "steerable headlights", etc.

Back-Up Cameras

What Is It?

Back-Up Cameras allow the driver to view the area behind the rear bumper and see small objects that may be obstructed by the vehicle's blind spots, or may not ordinarily be visible at all. Backover crashes account for a small number of overall crashes, but these events are much more likely to involve small children, and have a high likelihood of fatality. It is estimated that back-up cameras could reduce this type of accident by as much as 46%.

Back-up cameras would likely be useful to suburban or passengers van drivers who often back out of a driveway, or urban drivers who frequently parallel park. They would be especially useful to older drivers, who often lack the flexibility necessary to turn and thoroughly check the blind spot.

There are several different versions of back-up camera systems. Some systems simply provide a view from the back of the vehicle, while others pair this view with a sensor that warns (audible alarm) if an object is detected too close to the back of the vehicle. Other systems will even apply the brakes automatically to prevent a potential collision.

Just the Facts

Why Would I Use This Technology?

Backover accidents account for a fairly small percentage of total accidents, but they are more likely to lead to severe injury or death. They provide an easily accessible view of the back of the vehicle, and may warn the driver if a potential crash/collision is detected. The primary benefit of this technology is that it makes it **much easier to monitor a difficult to see area around the vehicle and take corrective action as a result.**

What Do Drivers Think?

Opinion on back-up cameras is very positive, with 80% of surveyed drivers agreeing that the technology improves their safety. 96% of respondents found the technology easy to use. Older drivers were more likely than younger drivers to be interested in the system. However, many drivers (67% in one survey) believe that the system will be active regardless of the speed traveled, when in fact it will not.

How Well Does It Work?

Backup cameras reduce the likelihood of a backover accident by at least 46%. It is difficult to say whether this extends to differences in insurance claims, as the likelihood of a backover crash is small to begin with.

Who Benefits Most?

Back-up cameras are currently installed in about 25% of all vehicles, and would likely be useful to suburban drivers who often back out of a driveway, or urban drivers who frequently parallel park. They would be especially useful to older drivers, who often lack the flexibility necessary to turn and thoroughly check the blind spot.

In What Situations Doesn't It Work?

Many back-up camera systems will turn off if the vehicle is traveling faster than a certain speed (6 MPH in many implementations). Drivers should remember to continue to check their rearview mirrors, as some drivers, may become overly reliant on the camera.

Mobility Significance

Back-Up Cameras increase the driver's field of view, and may be especially useful to older drivers who have trouble stretching to check the blind spot.

Not All Systems Are Alike

Back-Up Cameras can be found under several different configurations. Image quality will vary between different implementations and conditions. Some cameras overlay guidelines onto the video. Others are connected to a sensor that will warn the driver if a rear obstacle is getting too close, and some may even automatically slow down the vehicle.

Different Names, Same Idea

Back-Up Cameras can be found under a number of different names, including:

- Rear-View Camera

Forward Collision Warning (FCW)

What Is It?

Forward Collision Warning systems alert you when your vehicle is about to collide with another vehicle some distance ahead of yours. This is a very common type of collision that results in about 1.4 million crashes per year, or about a quarter of all collisions. Research has shown that Forward Collision Warning systems can substantially reduce the risk and severity of a crash.

Forward Collision Warning systems may be most helpful in alerting the driver to dangerous situations, helping him or her to respond more quickly as the need arises. The type of warning that the systems use will vary between vehicles; some use a flashing light, while others use an alarm sound or vibration.

Forward Collision Warning systems should not be confused with Forward Collision Mitigation systems. Warning systems simply warn the driver when a collision is likely, but do not automatically apply the brakes. It is also important to keep in mind that different vehicles have the ability to detect different kinds of crashes. Some vehicles will only sound the alarm if it is about to collide with another moving vehicle, for example.

<u>Just the Facts</u>

Why Would I Use This Technology?

Forward collision crashes, in which the front of a vehicle collides with another vehicle on the road, are very common, accounting for up to a quarter of all crashes. FCW systems detect when the vehicle may be about to collide with another object, and alert the driver to encourage corrective action. The primary benefit of this technology is that it **alerts drivers to dangerous situations and allows the driver to take action quickly.**

What Do Drivers Think?

A recent survey showed that only about 24% of respondents were aware of FCW technology. Of those, 60% said the technology was easy to use, and 47% considered it to be useful. A study that asked about drivers' satisfaction with FCW found that drivers were moderately satisfied with it.

How Well Does It Work?

Studies of real cars have shown that FCW systems can reduce rear-end collisions by about 10%. Insurance studies have also shown that owners of cars equipped with FCW have lower claim rates than owners who do not have FCW.

Who Benefits Most?

FCW is most commonly available on luxury and higher-end vehicles, but that is expected to change in the next few years. Since all drivers need some help monitoring their surroundings, all types of drivers are expected to benefit equally.

In What Situations Doesn't It Work?

FCW systems can be camera-based or radar-based. Camera systems can be obstructed by build-ups of ice or snow, are less accurate at night, and can sometimes be "blinded" by sunrise and sunset. Radar-based systems are less susceptible to the time of day, but can be affected by snow and ice.

Mobility Significance

FCW would be beneficial for inattentive drivers, allowing them to react more quickly to dangerous situations they might have missed.

Not All Systems Are Alike

FCW systems may be camera-based or radar-based, and have different weaknesses as a result (see "In What Situations Doesn't It Work?"). Some systems use a flashing light to indicate a possible crash, while others play an alarming sound. Most importantly, some systems only detect possible collisions with moving vehicles, whereas others work with both moving and stationary vehicles. Some FCW systems can also "prepare" the brake to make braking more effective.

Different Names, Same Idea

Forward Collision Warning can be found under a number of different names, including:

- Crash Imminent Warning
- Pre-crash Warning

Forward Collision Mitigation (FCM) / Braking

What Is It?

Forward Collision Mitigation systems detect how far and fast the vehicle in front of your may be moving, and automatically apply the brakes if the driver does not responds himself. Many drivers fail to notice when they are entering into a potential crash situation. Even when they do notice, many will fail to apply the brakes quickly enough. Forward Collision Mitigation systems work to reduce the chance of crashes, and reduce the severity of collisions when they occur. It is estimated that these systems could prevent up to 1.2 million crashes per year.

Forward Collision Mitigation systems would be most useful for inattentive drivers who have trouble monitoring their surroundings at all times. It would also be especially helpful to drivers who have trouble reacting quickly to unexpected events, such as older drivers or those with disabilities.

Forward Collision Mitigation should not be confused with Forward Collision Warning. A Mitigation system will both warn the driver and slow the vehicle, whereas a Warning system will only warn the driver. Additionally, some systems will only detect other moving vehicles or vehicles traveling at a minimum speed, while others will detect both moving and stationary vehicles.

<u>Just the Facts</u>

Why Would I Use This Technology?

Forward collision crashes, in which the front of a vehicle collides with another vehicle on the road, are very common, accounting for up to a quarter of all crashes. FCM systems detect when the vehicle may be about to collide with another object, and automatically slow the vehicle. The primary benefit of this technology is that it **alerts drivers to dangerous situations and takes pre-emptive action to avoid a crash**.

What Do Drivers Think?

A recent survey showed that only about 24% of respondents were aware of FCM technology. Of those, 60% said the technology was easy to use, and 47% considered it to be useful. A study that asked about drivers' satisfaction with FCM found that drivers were moderately satisfied with it. Drivers who experienced a "false braking event", in which the system detects a crash that isn't happening and stops the vehicle, tend to have a lower opinion of these systems.

How Well Does It Work?

The Insurance Institute for Highway Safety estimates that these systems could reduce crashes by up to 20%, preventing 66,000 serious crashes and 879 fatal crashes per year.

Who Benefits Most?

FCM systems would be most useful for inattentive drivers who have trouble monitoring their surroundings at all times. It would also be especially helpful to drivers who have trouble reacting quickly to unexpected events, such as older drivers or those with disabilities.

In What Situations Doesn't It Work?

Camera-based FCM systems are less effective than radar-based systems, as these do not work as well at night and can be "blinded" by sunrise and sunset. It is also important for the driver to remain vigilant, and not become too reliant on the system for warnings and help.

Mobility Significance

FCM would be beneficial for inattentive drivers, allowing them to react more quickly to dangerous situations they might have missed. It would be especially helpful for drivers who react more slowly to their surroundings, whether due to age or disability.

Not All Systems Are Alike

FCM systems may be camera-based or radar-based, and have different weaknesses as a result (see "In What Situations Doesn't It Work?"). In addition, not all systems are capable of detecting stationary vehicles, or slowing the vehicle at the same rate.

Different Names, Same Idea

Forward Collision Mitigation can be found under a number of different names, including:

- Crash Imminent Brake (CIB)
- Autonomous Emergency Braking (AEB)
- Emergency Brake Assist (EBA)
- Predictive Brake Assist (PBA)
- Pre-crash warning and braking systems (PCWBS)

Lane Departure Warning (LDW)

What Is It?

Lane Departure Warning systems alert you whenever you unintentionally drift too close to the edges of the lane. Approximately 1.6 million lane departure accidents occur each year, accounting for more than a quarter of all vehicle accidents. Lane Departure Warning has been shown to be effective in improving a driver's ability to control the vehicle, and research suggests that these systems could substantially reduce the risk of accident.

Lane Departure Warning is particularly helpful for drivers who do a lot of driving on highways and rural roadways. It is less helpful for city drivers. The warning type varies between car manufacturers; some use an alarm sound, while others cause the driver's steering wheel or seat to vibrate, creating a feeling like driving over a rumble strip.

Lane Departure Warning systems should not be confused with Lane Departure Prevention systems (sometimes referred to as lane keeping assistance). Warning systems provide a warning, but leave any corrective actions up to you. Prevention systems gently steer the car to automatically re-center you in the lane.

<u>Just the Facts</u>

Why Would I Use This Technology?

Lane departure accidents are one of the most common accident types, accounting for 1.6 million road accidents per year. Studies have estimated that Lane Departure Warning (LDW) systems could prevent up to 179,000 accidents per year, including 7,500 fatal crashes. The primary benefit of this technology is that it **reduces accidents**.

What Do Drivers Think?

Studies that gave drivers a LDW system to try found that 85% of drivers reported liking the system and found it to be useful. 93% used the system at least sometimes, 77% felt it increased their safety, and 80% would buy the system again; 67% of drivers felt that LDW made them safer drivers.

How Well Does It Work?

LDW systems have been shown to improve lane-keeping by up to 34%, and in one study cut unintentional lane crossing in half. Drivers who used LDW systems also became more likely to use their turn signals, especially if they drove often on highways. Some in the driving safety research community have questioned the extent to which many regular drivers actually leave LDW systems active, and thus are able to benefit from the warnings. A need for additional naturalistic observation of how drivers interact with such systems and specific implementations of warning systems has been suggested.

Who Benefits Most?

LDW systems are most useful to drivers who often drive on highways and rural roadways. Some research suggests that the systems may provide greater benefit to younger drivers.

In What Situations Doesn't It Work?

LDW systems work by "looking" at the road's lane markings. These systems are less accurate on roads where lane markings are in poor condition, or in cases where bad weather obscures the markings. They are also somewhat less accurate at night or other low-light situations. These systems are better suited to highway driving, and will be less useful in cities.

Mobility Significance

LDW systems may increase driver confidence, and also help drivers realize when they may not be paying sufficient attention to the roadway.

Not All Systems Are Alike

Some manufacturers put the LDW system on by default, while others leave it off by default. Remember, you won't benefit from the system if it's not turned on, so check your car's manual to be sure you how the LDW system is activated and are aware of any conditions that will impact its function. Some LDW systems use an alarm sound, while others make the steering wheel or seat vibrate, similar to the feeling of a rumble strip.

Different Names, Same Idea

Lane Departure Warning systems can be found under a number of different names. As noted earlier, lane departure warning systems are conceptually different from lane keeping assistance systems that actively intervene in vehicle steering to aid in keeping your vehicle within lane boundaries. The latter technologies are often identified by terms such as "lane keeping assistance" and "lane keeping support".

Appendix E: Initial Conceptualization of Possible Rating Factors Shared with Experts for Discussion and Comment

As part of the initial task set for the project, we attempted to develop an exhaustive list of potential factors that might impact the overall effectiveness / safety benefit of a given technology. The resulting conceptualization and factor listing was then shared and discussed with a number of identified experts across academic, governmental, and industry settings. The material in the next section represents a version of the document that was used as part of these discussions. (It evolved and was updated during the course of the sequential discussion with the various experts.)

In developing the listing of potential factors, our intention was to start with a list of all the variables that one would ideally like to consider in assessing a technology and then move to evaluate what factors were practical to consider based on available sources of objective data. It was assumed that a significant reduction in the number and form of factors that could realistically be used in developing an objectively based rating system would occur. The eventual reduction and simplification of the primary rating methodology can be observed in the Technology Review sheets included in this report and the method of rating the broad safety impact of a technology eventually proposed in this report. This earlier material is reproduced here as it does still have some conceptual relevance to the assessment of driver vehicle interfaces (DVIs) and driver involved safety systems at different levels of assessment.

Conceptualization of a Technology Safety Benefit Rating System

(Based on November 15, 2012 version with minor additions.)

Conceptual contributors to an in-vehicle technology (IVT) rating system – The following conceptual grouping has been developed as a starting point for our consideration of factors that may contribute to the overall benefits or costs of a particular in-vehicle technology. Potential factors are broken out in more detail starting in the next section.

- **Estimates of the significance of the safety area** - consideration of variables such as number of crashes, injuries, and fatalities that might be reduced through the use of technology. Measures might be drawn from statistical databases such as FARS and published statistical modeling research and other estimation methods. Identification and development of methodology for converting such information from this and other categories into meaningful ratings will be a key focus of the early portion of the project. While a full review of factors is necessary, elements of published crash reports such as location, e.g. intersection, highway, parking lots, rural road, etc., as well as types of collisions, e.g. lane departures, forward collision etc., and vehicle occupant information will need to be considered in weighting the benefits of a given technology.
- **Potential efficacy of the technology** – to what extent the technology might be expected to impact the identified area of safety concern. This domain focuses on how the technology can be expected to perform / benefit the drivers across all ages under ideal conditions. Potential benefits need to be weighed against a matrix variables in the categories below. Efficacy ratings also need to take into account the amount of learning required to derive maximum benefit. Some technologies do not require any driver involvement to provide benefit. Others require some understanding and experience with the technology. Thus ratings providing an estimate of benefits at both the novice and experienced user levels are likely to be a useful addition.
- **Potential drawbacks or limitations of the technology, including basic usability** – depending on how the evaluation matrix is eventually organized, this aspect is likely to involve the most sub-evaluations. Limitations of technologies include:
    - conditions under which the technology will not operate, performance may degrade, or actual failure may occur (dirt on camera lenses, weather, speed, tolerance boundaries);
    - extent to which the technology adds to the overall demand on the older driver (cognitive, visual, manipulative, auditory and haptic / tactile domain demands);
    - potential distraction that may under certain conditions introduce safety risks;
    - extent to which trust in the technology is required to derive benefit;
    - ease of learning how to effectively utilize the technology – which may involve the intuitiveness of the mental model, frequency of confusion experienced by drivers, effort or time (number of interactions) required to become proficient;
- **Potential connectivity and modularity of the technology** – although not a major consideration in the near-term, this factor will become increasingly important over the next few years. Consumers increasingly expect to be able to control their mobile devices

from their steering wheels, or conversely, to augment the vehicle with data from the mobile device (turn-by-turn navigation, for instance). Just as we ask to what extent an in-vehicle technology can integrate with external devices, we might also ask to what extent the technology can be separated from the vehicle. Is the technology available only as part of a high-end luxury package, or are there options that make the technology that make the technology more affordable and/or accessible to the average consumer. These factors may be thought of as "vehicle adjacent" contributors.

Potential Rating Factors

The following is a working initial list of rating factors that are potentially relevant to assessing the safety benefits or costs associated with in-vehicle technologies. This listing is in the process of being expanded and "detailed out". The current conceptual grouping below is seen as a starting point and may well change based on further thought and input from collaborators. Our intent is to develop this into an exhaustive list of theoretically relevant rating factors.

- **Safety Significance**
    - **Crash Reduction** (crashes / injury / fatality mitigation)
    - **Risk Reduction** (potential increased situational awareness benefits, i.e. back-up cameras, blind spot identification, lane departure, various warning systems)
- **Potential Efficacy**
    - **Ease of Learning** – This refers to the extent to which one has to learn how to use a system to benefit from the technology - which may involve the intuitiveness of the mental model, frequency of confusion experienced by drivers, effort or time (number of interactions) required to become proficient. Efficacy may vary to the extent to which a driver is a novice or experienced user of a system.
- **Drawbacks or Limitations**
    - **Limiting Conditions**
        - **Technical Limitations** – This refers to conditions under which the technology will not operate, performance may degrade, or actual failure may occur – i.e. weather, speed, and tolerance boundaries.
        - **Behavioral Adaptation -** This refers to unintended behaviors that can arise from repeated use of a new in-vehicle technology (IVT). For instance, studies on seatbelt use in the mid-90s showed that belted drivers drove at higher speed and gave themselves less headway between the forward vehicle. Similarly, younger drivers with access to rear-view cameras are much less likely to physically turn to look out of the rear-view window.

- **Added Demand**
  - **Cognitive Demand (Workload)**
  - **Visual Demand** - This refers to the amount of time that a driver must spend looking at an IVT in order to use it. The more time spent looking at something other than the road, the riskier a technology becomes. A number of experimental standards have been proposed for measuring visual demand.
  - **Manipulative (Motor) & Tactile Demand** - This refers to the complexity of physically manipulating the controls of an IVT. Technologies with simpler manipulative demands are less likely to distract the driver, whereas those that are more complicated or provide ambiguous tactile feedback are likely to be less safe and convenient to use. For example, a manual gearshift is simple to use, but using a stereo with dozens of smaller buttons presents much greater manipulative demand.
  - **Auditory Demand**
- **Distraction** – This refers to potential distraction that may arise from a technology that may under certain conditions introduce safety risks (some conceptual overlap with added demand – but not the same construct)
- **Trust** – This refers to the extent to which trust is required to derive benefit.
- **Ease of Learning** – *(This section is also listed under the primary heading, "Potential Efficacy"; there is a conceptual overlap here.)* This refers to the extent to which one has to learn how to use a system to benefit from the technology - which may involve the intuitiveness of the mental model, frequency of confusion experienced by drivers, effort or time (number of interactions) required to become proficient. Efficacy may vary to the extent to which a driver is a novice or experienced user of a system.

- <u>**Other categories for integration or breakout**</u>
  - **Frequency of Use** – This refers to the extent to which a system may have benefit if engaged, but data indicates a significant number of cases where drivers do not engage the technology for various reasons (not aware of how to engage, have actively disengaged, etc.) and thus tend not to derive benefit.
  - **Impact on Stress** - This can be conceptualized as a fairly direct measure of the driver's physiological comfort level, and can be measured by various physiological measures such as skin conductance and heart rate readings. (Self-report ratings are also relevant here but may or may not be seen as having as much objective validity – an open point for consideration.)
  - **Consumer Awareness, Use, and Satisfaction**
- **External Safety** - This refers to an in-vehicle technology's impact on the safety of persons or items outside the vehicle, such as pedestrians and intersections. Collision

detection systems that automatically slow the car have been shown to reduce the severity of pedestrian injuries in these types of crashes.

- **Comfort** – This refers to technologies largely identified as being outside of the primary safety domain but that may, nonetheless, provide some indirect safety benefits.
- **Convenience -** This refers to technologies largely identified as being outside of the primary safety domain but that may, nonetheless, provide some indirect safety benefits.

Detail on a Partial Listing of Rating Factors

From the list of potential rating factors, we intend to identify which factors are the most appropriate and feasible to consider in the near-term for a first generation rating system. One key criterion for identifying initially relevant factors is the availability of data to support the development of objective rating criteria. The following subsections cover a number (but not all) of the factor areas from the current list.

Crash Reduction

Data Sources

- Fatality Analysis Reporting System (FARS, NHTSA)
- Highway Loss Data Institute (HLDI, IIHS)
- National Automotive Sampling System (NASS, NHTSA)
- National Motor Vehicle Crash Causation Survey (NMVCSS, NHTSA)
- General Estimates System (GES, NHTSA)
- Consumer Reports safety data

Research Examples

- Aarts, L., & van Schagen, I. (2006). Driving speed and the risk of road crashes: A review. *Accident Analysis & Prevention*, *38*(2), 215–224. doi:10.1016/j.aap.2005.07.004
- Adell, E., Várhelyi, A., & Fontana, M. (2011). The effects of a driver assistance system for safe speed and safe distance–A real-life field study. *Transportation Research Part C: Emerging Technologies*, *19*(1), 145–155. doi:10.1016/j.trc.2010.04.006
- Cummings, P., & Grossman, D. C. (2007). Antilock brakes and the risk of driver injury in a crash: A case–control study. *Accident Analysis & Prevention*, *39*(5), 995–1000. doi:10.1016/j.aap.2007.01.005
- Curry, A.E., Hafetz, J., Kallan, M.J., Winston, F.K., Durbin, D.R. (2010). Prevalence of teen driver errors leading to serious motor vehicle crashes. *Accident Analysis & Prevention, 43(4),* 1285-1290. DOI:10.1016/j.aap.2010.10.019.
- Farmer, C. M. (2001). New evidence concerning fatal crashes of passenger vehicles before and after adding antilock braking systems. *Accident Analysis & Prevention*, 33(3), 361-369.

- Georai, A., Zimmermann, M., Lich, T., Blank, L., Kickler, N., & Marchthaler, R. (2009). New approach of accident benefit analysis for rear end collision avoidance and mitigation systems. *ENHANCED SAFETY OF VEHICLES.[vp]. 15-18 Jun*.
- Kusano, K. D., & Gabler, H. C. (2010). Potential Occupant Injury Reduction in Pre-Crash System Equipped Vehicles in the Striking Vehicle of Rear-end Crashes. *Annals of Advances in Automotive Medicine/Annual Scientific Conference*, *54*, 203.
- Page, Y., Foret-Bruno, J. Y., & Cuny, S. (2005). Are expected and observed effectiveness of emergency brake assist in preventing road injury accidents consistent. *Paper No 05*.
- Page, Y., Hermitte, T., & Cuny, S. (2011). How Safe is Vehicle Safety? The Contribution of Vehicle Technologies to the Reduction in Road Casualties in France from 2000 to 2010. *Annals of Advances in Automotive Medicine/Annual Scientific Conference*, *55*, 101.
- Suzuki, K., & Yamada, K. (2010). Method for Evaluating Effectiveness of Information Presentation in Terms of Collision Avoidance. *International Journal of Intelligent Transportation Systems Research*, *9*(1), 37–46. doi:10.1007/s13177-010-0023-8

Experimentally Derivable in Near-Term: Yes

Expert Opinion Available: Yes

Ease of Learning

Data Sources: again, difficult to quantify in a global sense. Data from Consumer Reports and JD Power, which assess convenience and ease of use, might make an acceptable proxy.

Research Examples

- Comte, S. L. (2000). New systems: new behaviour? *Transportation Research Part F: Traffic Psychology and Behaviour*, *3*(2), 95–111. doi:10.1016/S1369-8478(00)00019-X
- Llaneras, R. E. (2007). Safety Related Misconceptions and Self-Reported Behavioral Adaptations Associated With Advanced In-Vehicle Systems: Lessons Learned From Early Technology Adopters. *PROCEEDINGS of the Fourth International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design*.
- Warner, H. W., & Åberg, L. (2008). The long-term effects of an ISA speed-warning device on drivers' speeding behaviour. *Transportation Research Part F: Traffic Psychology*, *11(2)*, 96-107.

Experimentally Derivable in Near-Term: Yes

Expert Opinion Available: ?

Cognitive Demand (Workload)

Data Sources: comparing workload effects using something like the n-back or clock visualization tasks might prove useful.

Research Examples:

- Brookhuis, K. A., van Driel, C. J. G., Hof, T., van Arem, B., & Hoedemaeker, M. (2009). Driving with a Congestion Assistant; mental workload and acceptance. *Applied Ergonomics*, *40*(6), 1019–1025. doi:10.1016/j.apergo.2008.06.010
- Davidse, R. J., Hagenzieker, M. P., van Wolffelaar, P. C., & Brouwer, W. H. (2009). Effects of In-Car Support on Mental Workload and Driving Performance of Older Drivers. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *51*(4), 463–476. doi:10.1177/0018720809344977
- Ghazizadeh, M., & Boyle, L. N. (2010). Influence of Driver Distractions on the Likelihood of Rear-End, Angular, and Single-Vehicle Crashes in Missouri. *Transportation Research Record: Journal of the Transportation Research Board*, *2138*(-1), 1–5. doi:10.3141/2138-01
- Mehler, B., Reimer, B., Coughlin, J. F., & Dusek, J. A. (2010). Impact of Incremental Increases in Cognitive Workload on Physiological Arousal and Performance in Young Adult Drivers. *Transportation Research Record: Journal of the Transportation Research Board*, *2138*(-1), 6–12. doi:10.3141/2138-02
- Merat, N., & Jamson, A. H. (2008). The Effect of Stimulus Modality on Signal Detection: Implications for Assessing the Safety of In-Vehicle Technology. *Human Factors*, *50*(1), 145–158. doi:10.1518/001872008X250656
- Palinko, O., Kun, A. L., Shyrokov, A., & Heeman, P. (2010). Estimating cognitive load using remote eye tracking in a driving simulator. In *Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications* (pp. 141-144). ACM.

Experimentally Derivable in Near-Term: Yes

Expert Opinion Available: Yes

Visual Demand

Data Sources: ISO recommends a "visual occlusion" method to gauge the amount of pure visual demand imposed by an IVT. Eye tracking and manual coding of visual glance behavior has also been used extensively.

Research Examples

- Birrell, S. A., & Young, M. S. (2011). The impact of smart driving aids on driving performance and driver distraction. *Transportation Research Part F: Psychology and Behaviour*, *14*(6), 484–493. doi:10.1016/j.trf.2011.08.004
- Devonshire, J. M. (2012). Effects of Automotive Interior Lighting on Driver Vision. doi:10.1582/LEUKOS.2012.09.01.001
- Engström, J., Åberg, N., Johansson, E., & Hammarbäck, J. (2005). Comparison between visual and tactile signal detection tasks applied to the safety assessment of in-vehicle information systems. In *Proceedings of the Third International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design* (pp. 232-239).

- Klauer, S. G., Dingus, T. A., Neale, V. L., Sudweeks, J. D., & Ramsey, D. J. (2006). *The Impact of Driver Inattention On Near-Crash/Crash Risk* (No. DOT HS 810 594) (pp. 1–224).
- Neale, V. L., Dingus, T. A., Klauer, S. G., Sodweeks, J., & Goodman, M. (2005). An Overview of the 100-Car Naturalistic Study and Findings. *National Highway Traffic Safety Administration, Paper*, (05-0400).
- Pettitt, M. A. (2008). Visual demand evaluation methods for in-vehicle interfaces. (Doctoral dissertation, University of Nottingham).
- Stevens, A., Burnett, G., & Horberry, T. (2010). A reference level for assessing the acceptable visual demand of in-vehicle information systems. *Behaviour & Information Technology*, *29*(5), 527–540. doi:10.1080/01449291003624212
- Yee, S., Nguyen, L., Green, P., Oberholtzer, J., & Miller, B. (2007). Visual, auditory, cognitive, and psychomotor demands of real in-vehicle tasks. UMTRI Technical Report (UMTRI-2006-20).

Experimentally Derivable in Near-Term: Yes

Expert Opinion Available: Yes

Manipulative (Motor) & Tactile Demand

Data Sources: ?

Research Examples

- Engström, J., Åberg, N., Johansson, E., & Hammarbäck, J. (2005). Comparison between visual and tactile signal detection tasks applied to the safety assessment of in-vehicle information systems. In *Proceedings of the Third International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design* (pp. 232-239).
- Just, M. A., Carpenter, P. A., Keller, T. A., Emery, L., Zajac, H., & Thulborn, K. R. (2001). Interdependence of nonoverlapping cortical systems in dual cognitive tasks. *NeuroImage*, *14*(2), 417–426. doi:10.1006/nimg.2001.0826
- Merat, N., & Jamson, A. H. (2008). The Effect of Stimulus Modality on Signal Detection: Implications for Assessing the Safety of In-Vehicle Technology. *Human Factors*, *50*(1), 145–158. doi:10.1518/001872008X250656
- Yee, S., Nguyen, L., Green, P., Oberholtzer, J., & Miller, B. (2007). Visual, auditory, cognitive, and psychomotor demands of real in-vehicle tasks. UMTRI Technical Report (UMTRI-2006-20).

Experimentally Derivable in Near-Term: Yes

Expert Opinion Available: ?

Auditory Demand

Data Sources: no standardized data source.

Research Examples

- Kun, A., Paek, T., & Medenica, Z. (2007). The effect of speech interface accuracy on driving performance. *In INTERSPEECH (pp. 1326-1329).*
- Kun, A. L., Shyrokov, A., & Heeman, P. A. (2010). Spoken tasks for human-human experiments: towards in-car speech user interfaces for multi-threaded dialogue, In *Proceedings of the 2nd International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (pp. 57-63). ACM.
- Merat, N., & Jamson, A. H. (2008). The Effect of Stimulus Modality on Signal Detection: Implications for Assessing the Safety of In-Vehicle Technology. *Human Factors*, *50*(1), 145–158. doi:10.1518/001872008X250656
- Shyrokov, A. (2006). Setting-up experiments to test a multithreaded speech user interface. (Vol. 54). Technical report ECE.
- Yee, S., Nguyen, L., Green, P., Oberholtzer, J., & Miller, B. (2007). Visual, auditory, cognitive, and psychomotor demands of real in-vehicle tasks. UMTRI Technical Report (UMTRI-2006-20).

Experimentally Derivable in Near-Term: Yes

Expert Opinion Available: Yes

Behavior Adaptation

Data Sources: behavior adaptation is not represented in large databases. However, defining some standard metrics by which to measure behavior adaptation across multiple technologies (reaction time, vehicle telemetry changes, etc.) might make a promising side project.

Research Examples

- Adell, E., Várhelyi, A., & Fontana, M. (2011). The effects of a driver assistance system for safe speed and safe distance–A real-life field study. *Transportation Research Part C: Emerging Technologies*, *19*(1), 145–155. doi:10.1016/j.trc.2010.04.006
- Breyer, F., Blaschke, C., Farber, B., Freyer, J., & Limbacher, R. (2010). Negative Behavioral Adaptation to Lane-Keeping Assistance Systems. *IEEE Intelligent Transportation Systems Magazine*, *2*(2), 21–32. doi:10.1109/MITS.2010.938533
- Kiefer, R. J., & Hankey, J. M. (2008). Lane change behavior with a side blind zone alert system. *Accident Analysis & Prevention*, *40*(2), 683–690. doi:10.1016/j.aap.2007.09.018
- Llaneras, R. E. (2007). Safety Related Misconceptions and Self-Reported Behavioral Adaptations Associated With Advanced In-Vehicle Systems: Lessons Learned From Early Technology Adopters. *PROCEEDINGS of the Fourth International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design*.
- Rudin-Brown, C. M. (2010). "Intelligent" in-vehicle intelligent transport systems: limiting behavioural adaptation through adaptive design. *IET Intelligent Transport Systems*, *4*(4), 252. doi:10.1049/iet-its.2009.0151

- Warner, H. W., & Åberg, L. (2008). The long-term effects of an ISA speed-warning device on drivers' speeding behaviour. *Transportation Research Part F: Traffic Psychology, 11(2),* 96-107.

Experimentally Derivable in Near-Term: Yes

Expert Opinion Available: Yes

External Safety

Data Sources

- Databases that specifically denote pedestrian vs. vehicle collisions.
- Databases that provide estimates of total damages (IIHS, HLDI, etc)

Research Examples

- Devonshire, J. M. (2012). Effects of Automotive Interior Lighting on Driver Vision. doi:10.1582/LEUKOS.2012.09.01.001
- Farah, H., Koutsopoulos, H. N., Saifuzzaman, M., Kölbl, R., Fuchs, S., & Bankosegger, D. (2012). Evaluation of the effect of cooperative infrastructure-to-vehicle systems on driver behavior. *Transportation Research Part C: Emerging Technologies*, *21*(1), 42–56. doi:10.1016/j.trc.2011.08.006
- Helmer, T., Scullion, P., Samaha, R. R., Ebner, A., & Kates, R. (2011). Predicting the Injury Severity of Pedestrians in Frontal Vehicle Crashes based on Empirical, In-depth Accident Data. *International Journal of Intelligent Transportation Systems Research, 9*(3), 139–151. doi:10.1007/s13177-011-0036-y
- Oh, C., Kang, Y. S., & Youn, Y. (2009). Evaluation of a brake assistance system (BAS) using an injury severity prediction model for pedestrians. *International Journal of Automotive Technology*, *10*(5), 577–582. doi:10.1007/s12239−009−0067−4

Experimentally Derivable in Near-Term: Yes

Expert Opinion Available: Yes

Stress

Data Sources: Stress can be quantified directly as changes in heart rate or galvanic skin response. Stress can also be conceived of as a byproduct or covariate of the various demand factors outlined above.

Research Examples

- AgeLab citations
- Jenness, J. W., Lerner, N. D., Mazor, S. D., & Osberg, J. S. (2008a). *Use of Advanced In-Vehicle Technology By Young and Older Early Adopters: Survey Results on Headlamp Systems*. National Highway Transportation Safety Administration.

- Jenness, J. W., Lerner, N. D., Mazor, S. D., & Osberg, J. S. (2008b). *Use of Advanced In-Vehicle Technology by Young and Older Early Adopters. Survey Results on Navigation Systems*. National Highway Transportation Safety Administration.
- Jenness, J. W., Lerner, N. D., Mazor, S. D., & Osberg, J. S. (2008c). *Use of Advanced In-Vehicle Technology by Younger and Older Early Adopters. Selected Results From Five Technology Surveys*. National Highway Transportation Safety Administration.
- Jenness, J. W., Lerner, N. D., Mazor, S. D., Osberg, J. S., & Tefft, B. C. (2007). *Use of Advanced In-Vehicle Technology By Young and Older Early Adopters: Results on Sensor-Based Backing Systems and Rear-View Video Cameras*. National Highway Transportation Safety Administration.
- Jenness, J. W., Lerner, N. D., Mazor, S. D., Osberg, J. S., & Tefft, B. C. (2008d). *Use of Advanced In-Vehicle Technology By Young and Older Early Adopters: Survey Results on Adaptive Cruise Control Systems* (pp. 1–110). National Highway Transportation Safety Administration.

Experimentally Derivable in Near-Term: Yes (readily)

Expert Opinion Available: Yes

Consumer Awareness, Use, and Satisfaction

Data Sources:

- Consumer Reports
- JD Power (specifically the Automotive Performance, Execution, and Layout subscale)

Research Examples:

- Böhm, M., Fuchs, S., Pfliegl, R., & Kölbl, R. (2010). Driver Behavior and User Acceptance of Cooperative Systems Based on Infrastructure-to-Vehicle Communication. *Transportation Research Record: Journal of the Transportation Research Board*, *2129*(-1), 136–144. doi:10.3141/2129-16
- Brookhuis, K. A., van Driel, C. J. G., Hof, T., van Arem, B., & Hoedemaeker, M. (2009). Driving with a Congestion Assistant; mental workload and acceptance. *Applied Ergonomics*, *40*(6), 1019–1025. doi:10.1016/j.apergo.2008.06.010
- Farah, H., Koutsopoulos, H. N., Saifuzzaman, M., Kölbl, R., Fuchs, S., & Bankosegger, D. (2012). Evaluation of the effect of cooperative infrastructure-to-vehicle systems on driver behavior. *Transportation Research Part C: Emerging Technologies*, *21*(1), 42–56. doi:10.1016/j.trc.2011.08.006
- Jenness, J. W., Lerner, N. D., Mazor, S. D., & Osberg, J. S. (2008a). *Use of Advanced In-Vehicle Technology By Young and Older Early Adopters: Survey Results on Headlamp Systems*. National Highway Transportation Safety Administration.
- Jenness, J. W., Lerner, N. D., Mazor, S. D., & Osberg, J. S. (2008b). *Use of Advanced In-Vehicle Technology by Young and Older Early Adopters. Survey Results on Navigation Systems*. National Highway Transportation Safety Administration.

- Jenness, J. W., Lerner, N. D., Mazor, S. D., & Osberg, J. S. (2008c). *Use of Advanced In-Vehicle Technology by Younger and Older Early Adopters. Selected Results From Five Technology Surveys*. National Highway Transportation Safety Administration.
- Jenness, J. W., Lerner, N. D., Mazor, S. D., Osberg, J. S., & Tefft, B. C. (2007). *Use of Advanced In-Vehicle Technology By Young and Older Early Adopters: Results on Sensor-Based Backing Systems and Rear-View Video Cameras*. National Highway Transportation Safety Administration.
- Jenness, J. W., Lerner, N. D., Mazor, S. D., Osberg, J. S., & Tefft, B. C. (2008d). *Use of Advanced In-Vehicle Technology By Young and Older Early AdoptersSurvey Results on Adaptive Cruise Control Systems* (pp. 1–110). National Highway Transportation Safety Administration.
- Jiménez, F., Liang, Y., & Aparicio, F. (2012). Adapting ISA system warnings to enhance user acceptance. *Accident Analysis & Prevention*, *48*, 37–48. doi:10.1016/j.aap.2010.05.017
- Joshi, S., Bellet, T., Bodard, V., & Amditis, A. (2009). Perceptions of Risk and Control: Understanding Acceptance of Advanced Driver Assistance Systems. *Human-Computer Interaction–INTERACT 2009*, 524–527.
- Van Der Laan, J. D., Heino, A., & De Waard, D. (1997). A simple procedure for the assessment of acceptance of advanced transport telematics. *Transportation Research Part C: Emerging Technologies, 5(1),* 1-10.
- Warner, H. W., Özkan, T., & Lajunen, T. (2010). Drivers' propensity to have different types of intelligent speed adaptation installed in their cars. *Transportation Research Part F: Traffic Psychology and Behaviour*, *13*(3), 206–214. doi:10.1016/j.trf.2010.04.005

Experimentally Derivable in Near-Term: Yes (surveys)

Expert Opinion Available: Yes

Comfort

Data Sources: Consumer Reports and JD Power will likely have survey subscales available that address user comfort factors.

Research Examples

- Caberletti, L., Elfmann, K., Kummel, M., & Schierz, C. (2010). Influence of ambient lighting in a vehicle interior on the driver's perceptions. *Lighting Research and Technology*, *42*(3), 297–311. doi:10.1177/1477153510370554
- Coughlin, J. F., Reimer, B., & Mehler, B. (2011). Monitoring, managing, and motivating driver safety and well-being. *Pervasive Computing, IEEE*, *10*(3), 14–21. doi:10.1109/MPRV.2011.54
- Devonshire, J. M. (2012). Effects of Automotive Interior Lighting on Driver Vision. doi:10.1582/LEUKOS.2012.09.01.001

- Kyung, G., & Nussbaum, M. A. (2008). Driver sitting comfort and discomfort (part II): Relationships with and prediction from interface pressure. *International Journal of Industrial Ergonomics 38(5)*, 526-538.

Experimentally Derivable in Near-Term: Yes (surveys)

Expert Opinion Available: Yes

Appendix F: Initial Consultations with Selected Industry Experts & Observers

Early in the conceptual development of the rating scale (November 2012), we solicited input on a number of formative questions from a number of safety experts, academics, and industry professionals. They were provided with a copy of the conceptualization of a rating factor list (See previous Appendix) and encouraged to offer their thoughts on the strengths and weaknesses of different methodologies for developing a rating scale from their unique perspectives. Input was initially obtained through conference calls with members of the project staff that typically ranged between one and two hours each. Many of the conversations provided valuable insights on the project that helped shape and focus our thinking going forward. Summaries of key content from these conversations (in the form of a follow-up memo to the person interviewed) were drafted and sent to each participant for review and comment. The memos were edited based on feedback from the participants. A report on the first 8 interviews and resulting refinements in our listing of potential factors to consider in technology ratings (*Conceptualization of a Technology Safety Benefit Ratings System: Initial Conversations with Select Experts*) was sent to AAAFTS on December 2. This document was updated on December 14th. As noted below, three of the experts provided feedback on an anonymous basis as they spoke as individuals as opposed to formally representing their respective organizations.

Listing of Experts

Tom Baloga was the Vice President of engineering for BMW North America at the time of our call; he retired shortly thereafter. His main interests include intelligent in-vehicle systems, vehicle safety, and connected vehicle systems.

Azim Eskandarian is a Professor of Engineering and Applied Science at George Washington University, as well as the director of the Center for Intelligent Systems Research and the SEAS Transportation Safety and Security Program. His research focuses on intelligent in-vehicle systems, driver assistance, and collision avoidance technologies.

James Jenness is a Senior Research Scientist at Westat Inc., and has contributed to a large number of automotive safety projects. His work includes a NHTSA sponsored, extensive survey-based study of in-vehicle technology adoption by older drivers.

Neil Lerner is the Manager of Human Factors projects at Westat, Inc., and has done substantial work examining the effects of distraction on driving behavior, as well as drivers' perceptions of emerging in-vehicle technologies. He has collaborated with James Jenness on a number of projects.

Dan McGehee is an Adjunct Professor of Mechanical and Industrial Engineering at the University of Iowa Public Policy Center, as well as the director of the University's Human Factors and Vehicle Safety Research Program. His primary interests are in driver performance and behavior, interface design, and technology testing

Michael Perel is the former chief of the Human Factors/Engineering Integration Division at NHTSA (now retired).

Anonymous Number One has worked in the automotive industry for three decades, and currently oversees safety research for a consortium of automobile companies.

Anonymous Number Two is a research scientist with over thirty years of experience in both the automotive industry and academia, whose work on driver attention and behavior has helped set guidelines in the United States.

Anonymous Number Three is a senior engineer at a major automotive company whose work focuses on vehicle safety technologies. As part of his role, he is involved with international efforts in system testing and often acts as a liaison to regularity and other non-governmental traffic safety institutions.

Introduction Provided to Industry Experts

After an initial phone conversation with project lead Bryan Reimer, the following written introduction was provided to our initial select group of experts to provide a context for an extended conference call.

Developing a Consumer-Oriented Rating System for In-Vehicle Technologies

We have been tasked by the AAA Foundation for Traffic Safety with identifying and developing objective measures that can be used to construct a consumer-oriented rating system for in-vehicle technologies (IVTs). The resulting rating system should allow consumers to compare and contrast the effectiveness and efficacy of a wide range of IVTs in a manner that assists them in making informed purchasing decisions. In essence, if you have $2000 to spend, what are the most effective technologies for you to allocate these limited resources? It is our belief that effectively communicated information can educate consumers on the relative benefits of various IVTs and ensure that, where data exists, they are aware of safety advantages systems offer.

Rating the benefits or usefulness of a technology as much as possible on an objective basis presents a number of conceptual and methodological challenges, and we believe that the success of this undertaking will be greatly influenced by the breadth of perspective and depth of experience that can be taken into consideration in various ways in its development. Consequently, we have included as a fundamental component of this project, reaching out to a select group of knowledgeable individuals early in the process for input and comment.

As a starting point, we have chosen to conceptualize both IVTs and their underlying rating factors into the domains of safety, comfort, and convenience. Although we are primarily concerned with the safety impact of IVTs at the present time, comfort and convenience factors will also be considered insofar as they have the potential to improve the driving experience, reduce stress and demand on the driver, and thus ultimately influence roadway safety. This taxonomy is also forward-looking, and will potentially allow an expanded focus on rating more comfort- or convenience-oriented features, such as navigation and entertainment systems, in the future.

There are a number of ways to conceptualize safety and thus there are a broad range of factors that are potentially relevant to assessing the safety benefits or costs associated with IVTs. A core factor might be an estimation of the number of crashes / injuries / fatalities reduced based on the use of a given technology. Other factors that may exert both positive and negative influences include the usability of a system, learnability, understandability of when and how to engage or depend upon a system, behavioral adaptation to a technology, cognitive workload, distraction, tolerance to environmental conditions, etc. From a strategic standpoint, we believe that it makes sense to begin with as comprehensive a list of theoretically relevant factors as possible. While academic in nature, we believe the development of a comprehensive list will provide a more transparent basis on which to justify the selection of actual rating components (i.e. explicit reasons can be provided as to why various factors were or were not included). From that list, we intend to identify which factors are the most appropriate and feasible to consider in the near-term for a first generation rating system. One key criterion for identifying initially relevant factors is the availability of data to support the development of objective rating criteria. Therefore, we are also seeking input on data sources or bodies of existing research from government institutions, research institutes, or industry-affiliated organizations that may be particularly useful in assessing various IVTs on key factors. It is important to note that we also aim to highlight key factors that may be essential additions to a refined second generation system but for which data is not currently available; this process may aid in highlighting areas where future research activity would be valuable.

In brief then, we are reaching out to key experts: a) for input on additions to the initial broad factor list, b) for perspectives on the reduction of the theoretical list to key factors that are most promising for initial utilization in an objectively oriented assessment systems, c) suggestions regarding useful data sources, and d) thoughts on what one might see as major issues in the development of such a rating system. As one such expert, we are reaching out to you in specific to see to what extent you might have an interest in contributing to the further development of this project. Or, alternatively, we would very much appreciate it if you have any thoughts regarding someone you believe that we should consider contacting about this undertaking. Our intent with this brief introduction is really to provide just enough background on the project to hopefully lead to a follow-up conversation or other communication.

In due course we aim to make all aspects of this project public, informing industry and other key stake holders of our intent and seeking their input where appropriate. However, it is our intent to flesh out the model more fully with a limited set of individuals prior to moving in this direction. Given that we are contacting a limited number of individuals on a selective basis at this time, we would appreciate it if you would be willing to keep the substance of this communication relatively confidential for the time being. We are happy to acknowledge you as a direct contributor or as an anonymous source. We appreciate your time and consideration.

Appendix G: Information Requests and Interaction with Industry

The project was initially introduced to automotive industry through a briefing to The Alliance of Automobile Manufacturers Safety Policy Committee on January 17, 2013 and Global Automakers Safety Committee on February 1, 2013. Materials from these briefings appear below. At the conclusion of the meetings we asked companies to supply a key contact, help in selecting technologies for rating, and relevant data to support the objective rating of technologies. Through contacts provide by attendees of these briefings as well as other industrial contacts, we assembled a list of contacts in the following vehicle manufactures: BMW, Daimler, Ford, General Motors, Honda, Jaguar Land Rover, Mitsubishi, Nissan, Subaru, Toyota, Volkswagen, and Volvo. Follow on discussions with contacts were held on March 5th 2013 and April 9th 2013 (see materials below). A request for information on technologies selected for rating was issued in early May. An introductory email and detailed questionnaire (long form) was then sent to the key contacts at all of the vehicle manufactures trying to gather detailed information to support the objective rating of technologies. At the request of several manufactures, we created a shorter questionnaire (short form) (see materials below). In addition to efforts focused on gathering data from vehicle manufacturers, we reached out contacts at Tier 1 suppliers including: Bosch, Continental, Delphi, Denso, Johnson Controls, Takata, Valeo, and Visteon, with request for information on the technologies.

We received considerable information on different technologies from a number of vehicle manufacturers and suppliers. Many of the contacts provided marketing information on various systems and a select number of contacts provided more specific technical detail. Numerous one-on-one conversations were held with manufacturers and suppliers to draw further insight on the availability of data suitable for the objective rating of technologies. Finally, on December 18th 2013, a briefing on the developed system was provided to solicit any final comments from industry supporters.

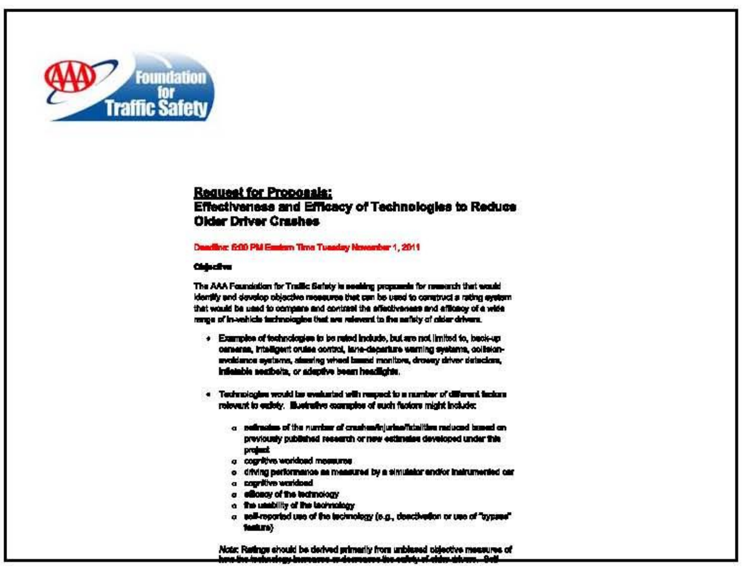
Formal information requests to industry included the materials listed below.

- PowerPoint slides that were used during our initial briefing of industry at the offices of the Alliance of Automotive Manufacturers and the Association of Global Automakers in Washington D.C.
- A supplemental document was provided to individuals who attended the aforementioned briefing to provide additional background on the project that attendees could share with colleagues to identify an appropriate representative or representatives who might serve as formal contacts for further interaction.
- Invitation to participate in industry discussion on rating project that took place on March 5th, 2013 – sent to contacts developed out of presentations before the Alliance of Automotive Manufacturers and the Association of Global Automakers as well as other contacts.
- Invitation to participate in follow-up industry discussion and presentation of concept materials that took place on April 9th, 2013.
- Cover letter / e-mail introduction to information request packet.
- A two page project introduction document "Project Synopsis". (Not included here.)
- A technology prioritization list soliciting input from industry on technologies they felt were most import to include in the rating project.

- A long form for providing detailed information. Individual forms were provided for each of the technologies selected for inclusion in the phase I evaluations along with a blank form that could be completed for any technology that industry representatives felt we should seriously consider including in the first round ratings.
- A "short" version of the form was subsequently developed to encourage sources that had limited time / resources available to contribute to the project.
- Also provided were early draft versions of the materials being developed for electronic stability control to provide additional context for the type of information we were attempting to develop on the other technologies. (Not included here.)

Briefing Slides

The following slides were used during our initial briefing of industry at the offices of the Alliance of Automotive Manufacturers and the Association of Global Automakers.



Foundation for Traffic Safety

Request for Proposals:
Effectiveness and Efficacy of Technologies to Reduce Older Driver Crashes

Deadline: 5:00 PM Eastern Time Tuesday November 1, 2011

Objective

The AAA Foundation for Traffic Safety is seeking proposals for research that would identify and develop objective measures that can be used to construct a rating system that would be used to compare and contrast the effectiveness and efficacy of a wide range of in-vehicle technologies that are relevant to the safety of older drivers.

- Examples of technologies to be rated include, but are not limited to, back-up cameras, intelligent cruise control, lane-departure warning systems, collision-avoidance systems, steering wheel based monitors, drowsy driver detectors, inflatable seatbelts, or adaptive beam headlights.
- Technologies would be evaluated with respect to a number of different factors relevant to safety. Illustrative examples of such factors might include:
  - estimates of the number of crashes/injuries/fatalities reduced based on previously published research or new estimates developed under this project
  - cognitive workload measures
  - driving performance as measured by a simulator and/or instrumented car
  - cognitive workload
  - efficacy of the technology
  - the usability of the technology
  - self-reported use of the technology (e.g., deactivation or use of "bypass" feature)

Note: Ratings should be derived primarily from unbiased objective measures of

Correspondence > Bryan Reimer, Ph.D. > (617) 452 – 2177 > reimer@mit.edu © 2013 MIT AGELAB 1

MIT AgeLab

Human Behavior Across the Lifespan

Our lab, located within the Engineering Systems Division at MIT, focuses on multi-disciplinary approaches to understanding and optimizing systems with a specific emphasis on studying issues across the life-span. This work falls into a number of domains:

- Human Factors
- Cognitive Engineering
- Neuroergonomics
- Psychology
- Psychophysiology
- Social Science
- Computer Science
- Ergonomics
- Computer Human Interface Design
- Augmented Cognition

MIT ESD

Correspondence > Bryan Reimer, Ph.D. > (617) 452 - 2177 > reimer@mit.edu

MIT AgeLab

Human Behavior Across the Lifespan

We have extensive history working with the automotive industry, engaging with:

- Audi
- BMW
- Chrysler
- Daimler
- Delphi
- Denso
- Fiat
- Ford
- General Motors
- Honda
- Hyundai
- Nissan
- Peugeot
- Takata
- Toyota
- Volkswagen
- Volvo
- ...and others

Correspondence > Bryan Reimer, Ph.D. > (617) 452 - 2177 > reimer@mit.edu

Correspondence > Bryan Reimer, Ph.D. > (617) 452 - 2177 > reimer@mit.edu

Vehicle Ratings

- Vehicle Crashworthiness
  - NHTSA
  - IIHS
- Overall Vehicle Experience/Satisfaction
  - Consumer Reports
  - Edmunds
  - Kelley Blue Book

These organizations rate the vehicle **holistically**, but do not rate individual technologies.

Emerging Technologies

Correspondence > Bryan Reimer, Ph.D. > (617) 452 – 2177 > reimer@mit.edu © 2013 MIT AGELAB

Correspondence > Bryan Reimer, Ph.D. > (617) 452 – 2177 > reimer@mit.edu © 2013 MIT AGELAB

Obstacles to Purchase & Adoption

1. Consumers have a **poor understanding** of how new in-vehicle technologies work, and what benefits they may have. (Llaneras 2007)

2. Experienced drivers may be **less inclined** to try unfamiliar technologies or to be receptive to their potential benefits. (Jenness et al. 2008, AAA Foundation for Traffic Safety et al. 2008)

Correspondence > Bryan Reimer, Ph.D. > (617) 452 – 2177 > reimer@mit.edu

Obstacles to Purchase & Adoption

3. Automobile sales **dropped 41%** during the Great Recession, and are only now approaching pre-Recession levels. (Alliance Sales Data 2012)

4. Consumers have more options than ever, but a much smaller amount of money with which to pursue them.

5. Consumers want to know which in-vehicle technologies make the most sense to invest in given their individual **driving needs and limited resources**.

Correspondence > Bryan Reimer, Ph.D. > (617) 452 – 2177 > reimer@mit.edu
© 2013 MIT AGELAB

Our Task

The AAA Foundation for Traffic Safety tasked us with creating a **data-driven rating system** for new in-vehicle technologies, analogous to NCAP crashworthiness, but extended to scalar ratings of individual technologies.

This system has the potential to educate and guide consumers towards more confident and strategic **purchasing decisions** that will enhance automotive safety.

Correspondence > Bryan Reimer, Ph.D. > (617) 452 - 2177 > reimer@mit.edu

Trusted Information Sources

MIT + AAA Foundation for Traffic Safety → AAA

Objective Research

Smart, Safe Consumer Advocate

54 Million Members

Correspondence > Bryan Reimer, Ph.D. > (617) 452 - 2177 > reimer@mit.edu

Conceptual Vision

We view the system's rating factors as falling into **three broad areas**.

The initial selection of technologies to be rated are oriented toward safety, but later versions could encompass other in-vehicle technologies.

Correspondence > Bryan Reimer, Ph.D. > (617) 452 – 2177 > reimer@mit.edu

Current Progress

Consulted a diverse set of experts during the identification of factors for the rating scale:

- Engineering/Human Factors specialists
  (industry)
- Safety specialists
  (government, foundations, industry & academia)
- Behavior/Cognition specialists
  (industry & academia)
- Public policy & marketing experts
  (industry, foundations & academia)

Correspondence > Bryan Reimer, Ph.D. > (617) 452 – 2177 > reimer@mit.edu
© 2013 MIT AGELAB

Current Progress

Rating Factors

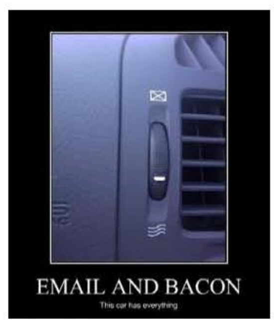
- Safety Significance
  - Crash Reduction
  - Risk Reduction
- Potential Efficacy
  - Ease of Learning
- Drawbacks & Limitations
  - Technical Limitations
  - Behavioral Adaptation
- Added Demand
  - Cognitive Workload
  - Visual Demand
  - Auditory Demand
  - Tactile Demand
- Distraction
- Stress
- Consumer Factors
  - Trust
  - Awareness
  - Frequency of Use
  - Satisfaction

Correspondence > Bryan Reimer, Ph.D. > (617) 452 – 2177 > reimer@mit.edu

Current Progress

Technologies of Interest

- Electronic Stability Control?
- Lane Departure Warning?
- Lane Keeping Assist?
- Adaptive Cruise Control?
- Forward Collision Mitigation?
- Forward Collision Warning?
- Blind Spot Detection?
- Back-Up Cameras?



Correspondence > Bryan Reimer, Ph.D. > (617) 452 – 2177 > reimer@mit.edu
© 2013 MIT AGELAB

Correspondence > Bryan Reimer, Ph.D. > (617) 452 – 2177 > reimer@mit.edu
© 2013 MIT AGELAB

Timeline

- Next six months
  - Identify sources of existing data to complete evaluations of candidate technologies, in collaboration with expert panel.
- End of 2013
  - A rating system with ratings completed for five example technologies due to AAAFTS
- Start of 2014
  - AAAFTS / AAA has the option of publicizing the scale
- Beyond
  - Refinement of rating system through increased emphasis on empirical data

Correspondence > Bryan Reimer, Ph.D. > (617) 452 – 2177 > reimer@mit.edu

Going Forward

What Are We Looking For?

Support from your organization in the development and refinement of the rating system, specifically:

- A **contact** in your organization to facilitate communication and collaboration.
- Thoughts on **technologies** that you feel are important to include.
- Relevant **data** to support objectivity of the ratings.

Correspondence > Bryan Reimer, Ph.D. > (617) 452 – 2177 > reimer@mit.edu
© 2013 MIT AGELAB

We Need Data

- This rating system is to be data-driven. Therefore, we see value in considering:
  - Internal **studies** on in-vehicle technologies
  - **Specification** documents for technologies
  - Relevant **research** on:
    - The driver-vehicle relationship
    - How the technology operates, and in what conditions
    - Crash reduction potential
- **Shared data can be kept de-identified and confidential, if desired.**

Correspondence > Bryan Reimer, Ph.D. > (617) 452 – 2177 > reimer@mit.edu

Help Us Help You!

We view automotive technology companies as **key collaborators and contributors** to this project.

The more input that industry has on this rating system, the more **informative** it will be.

Ultimately, the rating system is a tool for **promoting technologies and educating consumers** that will enhance safety and confer maximal benefit to industry, government, and automotive safety.

Correspondence > Bryan Reimer, Ph.D. > (617) 452 – 2177 > reimer@mit.edu
©2013 MIT AGELAB

Correspondence > Bryan Reimer, Ph.D. > (617) 452 – 2177 > reimer@mit.edu
© 2013 MIT AGELAB

Summary

- We are developing a rating system for new in-vehicle technologies.
- In the near-term (2013), we will base ratings on currently available empirical data and the aggregate opinions of an expert panel.
- In the long-term, the goal is to base the ratings primarily on empirical-derived data.
- The overall focus is on helping consumers make informed purchasing decisions.

Our Task

Our ultimate goal is to produce a succinct rating system for new in-vehicle technologies that can be presented to consumers in a compact, understandable, and actionable manner.

Correspondence > Bryan Reimer, Ph.D. > (617) 452 – 2177 > reimer@mit.edu ©2013 MIT AGELAB

AGELAB
Life Tomorrow

Presentation Supplement 2013-3

Developing a Rating System for In-Vehicle Technologies

Bryan Reimer, Bruce Mehler,
Jonathan Dobres, & Joseph Coughlin

January 17, 2013

AgeLab: More than Just Age

The Massachusetts Institute of Technology AgeLab examines issues pertaining to system optimization and human behavior *across the lifespan*. Our work encompasses a range of domains including human factors, ergonomics, computer science, social science, cognitive engineering, and human-computer interface design.

Over the last several years, we have engaged with a large number of automotive manufacturers and suppliers to investigate behaviors, cognition, and physiological responses of drivers as they are interact with a variety of technologies, interfaces, and operating conditions. Laboratory research has included developing new methodologies for assessing driver distraction associated with visual-manipulative and cognitive demands, devising strategies for optimizing the driver-vehicle interface, and measuring the stress/workload of drivers' engagement with Advanced Driver Assistance Systems (ADAs).

Emerging In–Vehicle Technologies

In the last fifteen years, a large variety of new comfort, convenience and safety technologies have been introduced into the motor vehicle. However, publically available data is limited on the extent to which there are barriers to the effective use of these technologies. If aspects of a technology's design are hindering or preventing its adoption or use, its potential safety benefits are negated.

While consumer-focused resources that rate or assess the whole vehicle have existed for decades (crashworthiness tests, Consumer Reports, Edmunds, etc.), there is no comparable resource that rates these new in-vehicle technologies themselves. While newer programs like NCAP take note of the presence of a technology, they do not rate that technology *per se*.

Consumers often lack a strong understanding of the benefits of emerging in-vehicle technologies (Llaneras 2007). Combined with a change in attitude brought on by the financial constraints of the Great Recession, many consumers want to know which in-vehicle technologies make the most sense to invest in given their individual driving needs and limited resources. Moreover, as US demographics begin to shift older, it will be increasingly important to understand how this population segment understands and uses (or do does not use) new technologies (Coughlin & Reimer 2006).

Our Task

The AAA Foundation for Traffic Safety (AAA-FTS) tasked us with creating a data-driven rating system for new in-vehicle technologies. Conceptually similar in some respects to NCAP, this system has the potential to educate and guide consumers towards more confident and strategic purchasing decisions that will ultimately enhance automotive safety.

We conceive the rating system as one in which rating factors are classified into the broad areas of safety, comfort, and convenience. In the near-term we will emphasize the safety area, but this overlapping model gives us the flexibility to assess any in-vehicle technology.

In consultation with a variety of experts from government institutes, academia, and industry, we have developed a matrix of possible rating factors (e.g., accident reduction potential, ease of learning, cognitive workload, trust, etc.).

Going Forward

Over the next six months, we will assemble an expert panel to construct an initial set of ratings for a handful of in-vehicle technologies. A draft of the rating system is due to AAA-FTS by the end of 2013, at which time they will have the option of publicizing the results and developed methodologies. Our (MIT and AAA-FTS) intent is to then further refine the system through an increased emphasis on empirical data.

We believe that successful development of this rating system will be of great benefit to the industry, as it is will encourage consumers to consider purchasing technologies that are strategically useful for them. We are seeking support from your organization in the development and refinement of this rating system. In specific, we are looking for a contact in your organization to facilitate communication, thoughts on important technologies to include, and relevant data to support the objectivity of the ratings. Data might include internal studies, specification documents, and other relevant research on an in-vehicle technology that contributes to an objective evaluation of the nature of the benefit and effectiveness. Any data contributed to this project can be kept confidential at the request of the contributor.

We view automotive technology companies as key collaborators and contributors to this project. The more input that industry provides in the development on this rating system, the more informative it will be. Ultimately, we see the rating system as a tool for promoting useful technologies and educating consumers that will enhance safety and confer maximal benefit to industry, government, and automotive safety.

References

Coughlin, J., & Reimer, B. (2006). New Demands from an Older Population: An Integrated Approach to Defining the Future of Older Driver Safety (pp. 1–8). Presented at SAE International.

Llaneras, R. E. (2007). Safety Related Misconceptions and Self-Reported Behavioral Adaptations Associated With Advanced In-Vehicle Systems: Lessons Learned From Early Technology Adopters. *PROCEEDINGS of the Fourth International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design.*

About the Researchers

Bryan Reimer, Ph.D.

Bryan Reimer is a Research Engineer in the Massachusetts Institute of Technology AgeLab and the Associate Director of the New England University Transportation Center. His research seeks to develop new models and methodologies to measure and understand human behavior in dynamic environments utilizing physiological signals, visual behavior monitoring, and overall performance measures. Dr. Reimer leads a multidisciplinary team of researchers and students focused on understanding how drivers respond to the increasing complexity of the operating environment and on finding solutions to the next generation of human factors challenges associated with distracted driving, automation and other in-vehicle technologies. He directs work focused on how drivers across the lifespan are affected by in-vehicle interfaces, safety systems, portable technologies, different types and levels of cognitive load. Dr. Reimer is an author on over 70 peer reviewed journal and conference papers. Dr. Reimer is a graduate of the University of Rhode Island with a Ph.D. in Industrial and Manufacturing Engineering.

reimer@mit.edu
(617) 452-2177
http://web.mit.edu/reimer/www/

Bruce Mehler, M.A.

Bruce Mehler is a Research Scientist in the Massachusetts Institute of Technology AgeLab and the New England University Transportation Center, and is the former Director of Applications & Development at NeuroDyne Medical Corporation. He has an extensive background in the development and application of non-invasive physiological monitoring technologies and research interests in workload assessment, individual differences in response to cognitive demand and stress, and in how individuals adapt to new technologies. Mr. Mehler is an author on numerous peer reviewed journal and conference papers. He received an MA in Psychology from Boston University and a BS degree from the University of Washington.

bmehler@mit.edu
(617) 253-3534
http://agelab.mit.edu/bruce-mehler

Jonathan Dobres, Ph.D.

Jonathan Dobres is a postdoctoral research associate in the Massachusetts Institute of Technology AgeLab. Dr. Dobres's research interests are varied and include human-computer interaction, user experience design, visual attention, and visual learning. He received a BA, MA, and PhD in Psychology (Brain, Behavior, and Cognition) from Boston University. His research concerned the effects of feedback, or knowledge of results, on how people learn visual tasks, as well as computational approaches to visualizing changes in human perception. He has also worked for the Traumatic Brain Injury Model System at Spaulding Rehabilitation Hospital, part of a long-term national study on the effects of traumatic brain injuries. His current research focuses on driver behavior, and how interacting with in-vehicle technologies impacts driver behavior, cognition, and physiology.

jdobres@mit.edu
(617) 253 – 7728

Joseph F. Coughlin, Ph.D.

Joseph F. Coughlin is founder and Director of the Massachusetts Institute of Technology AgeLab and Director of the US Department of Transportation's Region I New England University Transportation Center. He served as the Chair of the Organization for Economic Cooperation & Developments 21-nation Task Force on Technology and Transportation for Older Persons, is a member of the National Research Council's Transportation Research Board Advisory Committee on the Safe Mobility of Older Persons. He served as a Presidential appointee to the White House Conference on Aging and has consulted or served on technology and design boards for BMW, Daimler, Nissan, and Toyota. Prior to joining MIT, Dr. Coughlin led the transportation technical services consulting practice for EG&G, a global Fortune 1000 science and technology firm.

coughlin@mit.edu
(617) 253-4978
http://www.josephcoughlin.com/

About the AgeLab

The AgeLab is a multi-disciplinary research center dedicated to improving quality of life for older adults and those who care about them. Based within the Engineering Systems Division at Massachusetts Institute of Technology, the AgeLab is uniquely suited to translate cutting edge scientific and technological breakthroughs into innovative solutions that help address challenges posed by the world's aging population.

The AgeLab views longevity as an opportunity to innovate – to invent a new definition of quality living throughout the lifespan. AgeLab activities set agendas of government and business, serve as a catalyst for change, and act as platforms to create new ways to remain engaged, connected, independent, and healthy.

Funded by businesses around the world, AgeLab research focuses on transportation, health & wellness, caregiving, longevity planning, shopping, lifelong engagement, and even play. AgeLab research informs the design of new technologies, aids in government policy decisions in the United States and abroad, and educates older adults and their families on important consumer issues.

MIT AGELAB

Contact Information

Agelab
Massachusetts Institute of Technology
77 Massachusetts Avenue
Room E40-291
Cambridge, MA 02139-4307
Phone: (617) 253-0753
Fax: (617) 258-7570
Email: agelab@mit.edu
Website: http://agelab.mit.edu

Invitation to Participate in Industry Discussion on Rating Project

Invitation / Agenda for March 5th Industry Discussion and Presentation Material

Dear [ ],

Thank you for helping to coordinate [COMPANY]'s collaboration with the MIT AgeLab on the AAA-FTS sponsored project to create a rating system for in-vehicle technologies. Our next step in this process will be to arrange a conference call among the industry representatives who have so far committed their company's support to the project. This call will be centered on discussing:

A proposed set of factors being considered for developing the underlying basis of the rating system. Factors currently being considered include the technology's accident reduction/prevention potential, frequency of use, visual, auditory and/or tactile demands, ease of learning, consumer trust and awareness, potential drawbacks and failure conditions, etc. The factors' relevance to the rating system and availability of related data should be taken into consideration. Input on additions / exclusions will be actively taken under advisement as will thoughts on concepts for translating the factors into system ratings.

Technologies for initial consideration in the rating system. Our mandate is to include a minimum of 5 technologies in the initial rating. Again, the availability of data for rating a given technology should be kept in mind when discussing this point. A general discussion of what data is (and is not) available for vehicle technologies and/or rating factors is intended. However, more specific discussions will be held privately with each manufacturer.

The level of depth the rating system should go into on a specific technology, i.e.. does FCW based upon a radar or vision system need to be discussed separately? What are manufacturers' thoughts on including with the ratings a listing of vehicle models that are available with the selected technologies?

What operating conditions (urban, highway) or user types (older drivers) should be considered for categorical grouping in the rating system? Our current view is that various technologies will show differential advantages depending on various user characteristics.

In the near-term, we plan to reduce the available data to a set of more digestible statistics, and then provide this information to a panel of experts, who will provide feedback that will ultimately inform the first version of the rating system. The panel will include leading figures from industry, academia, and government institutions. We are also seeking your input on potential additional panel members.

Other suggestions, concerns and thoughts from an industry perspective.

Next steps following the call - individual conversations with OEMs on specific technologies, follow-up with selected suppliers, and others.

We would like to schedule the teleconference within the next few weeks, ideally for March 5th, 5th, or 12th (9am – 5pm EST). Please let us know if you would be available for a call (approximately 2 hours long) on these dates.

Given the difficulty involved with scheduling this call across multiple parties, we would greatly appreciate a response on preferable data's and times as quickly as possible.

Bryan Reimer, Jonathan Dobres, and Bruce Mehler

MIT AgeLab

Invitation to Participate in Follow-Up Industry Discussion

Invitation / Agenda for April 9th Industry Discussion and Presentation Material

Everyone,

Attached you'll find a set of documents we'll be discussing during tomorrow's conference call. These documents incorporate feedback from our previous teleconference, as well as feedback from AAA-FTS and further elaboration on our part. Attached:

1. Rating Documentation: Explains the models and rationale that underlie the four broad rating categories that we envision presenting to consumers.

2. Consumer Technology Explanation: This documentation is intended to help you collaborate with us. It provides written explanations of what we see as useful in constructing public-facing descriptions of various in-vehicle technologies.

3. Electronic Stability Control: Example consumer-facing document. (This example is for illustrative purposes and is not intended to represent a fully developed presentation.)

4. Technologies: A listing of in-vehicle technologies that were discussed at the last conference call, with those that are currently prioritized for the first phase of the rating system highlighted.

As we emphasize in the documents themselves, while we consider the concepts as presented to be fairly well developed, we are open to input regarding additions, refinements, modifications, etc. that might improve the proposed approach. We are sharing these working documents with you specifically for the purpose of gathering input and your assistance in improving upon them. Over the course of tomorrows call our agenda is as follows:

1. Discussion of the proposed approach for calculating technology rating

2. Discussion of the proposed consumer focused technology explanation

3. Review of the ESC illustration

4. A discussion of what type of factors should be added to describing the theoretical limits of system

5. A discussion of what type of factors should be added to describing the human factors limitations involved with a system

6. A review of the in-vehicle technologies that are being considered for inclusion in the first phase of this project

7. Requests for information and supporting materials that may be useful in rating and explaining selected technologies

8. A survey of what theoretical and human factors elements are most crucial for describing each technology (identification of research gaps)

9. Supplier contacts that can help with #7 and #8

10. Next steps

We are looking forward to the conversation tomorrow.

Best wishes.

Bryan, Bruce and Jon

Cover Letter

Industry Input to the MIT AgeLab AAA-FTS In-Vehicle Technology Rating Project

May 3, 2013

Dear Colleague:

Attached with the e-mail you will find the following:

- **Technology Information Forms** – A form for each of seven technologies. The cover page explains the overall purpose of the form. They are being provided in Word and Google Doc formats so that input can be typed directly into a document. Alternatively, a link is provided below for an on-line version. While sending material back to us in document format has some advantages at our end, the on-line version is also being offered since it provides a totally anonymous method of providing input since an e-mail connection is not involved.

  https://docs.google.com/forms/d/1Uqq9fpQ1fiC4_GWtMMtTmzfvQK0i2al9a-0taaWq2mg/viewform?sid=7f733369c89be39&token=jafPaz4BAAA.lWHXDm1-r0D0CH7cxn9GkQ.ZHkDTNa9RIw1KDQRVrR5SQ

  You are equally welcome to submit one form per technology or to distribute multiple copies to various members of your organization to fill-in selected portions or to obtain a broader perspective from within the organization.

- **Technology Prioritization List** – Please use this form to give us feedback on technologies that you feel should be included in this consumer education project and the relative priority with which they should be considered.

- **Technology Information Form (blank)** – If you feel strongly that there are additional technologies that should be included in the first round of the evaluation project, you are welcome to use the blank form to write-in the technology name, fill-in information on the technology, and submit.

- **Electronic Stability Control** (Illustrative Example) – ESC is being included as a reference technology for which there is a relatively extensive body of objective information and research available. This brief document is an early "work in progress" example that is intended to give an idea of the kind of information we are seeking. This is meant to stimulate thought and should not in any way constrain the kinds of information that you feel may be useful to contribute to inform the public and the evaluation process.

We realize that a great deal of effort could be invested in exhaustively completing these forms, but also believe that this project provides you, as representatives of the automotive industry, with an opportunity to contribute to educating the public about the potential value of investing in various safety technologies. Your experience, insight, and any data sources that you can identify that contribute to the development of a better understanding of safety benefits of these technologies is greatly appreciated. Please don't hesitate to contact us with any questions.

Bryan Reimer, Bruce Mehler, & Jonathan Dobres

Technology Prioritization Rating Form

Technologies

For the first round of technology ratings, we are currently considering the technologies listed immediately below. (Note: *Electronic Stability Control* is being included specifically as a reference technology for which there is a relatively extensive body of objective information and research available.)

- Lane Departure Warning
- Back-up Cameras
- Forward Collision Warning
- Forward Collision Mitigation
- Adaptive / Smart Cruise Control (headway management)
- Adaptive Headlamps
- Electronic Stability Control

Other Technologies under Consideration:

Active input is being sought from industry and other sources to prioritize other technologies to be included in the assessment process. Please help us in this process by adding additional technologies to this list that you feel should be considered. Then please rank the list in terms what you see as their order of importance. If you feel a technology should not be included, mark with an "x" or cross out.

__ Active rollover protection

__ Automated/Assisted Parking

__ Anti-lock Braking Systems

__ Blind Spot Detection

__ Driver Monitoring Systems

__ Emergency Brake Assist

__ Lane keeping aids

__ Night Vision/Low-light Systems

__ Three-point seatbelts for all positions

__ Active seat belts

__ Seat belt reminders

__ Active head restraints

__ Rear crash warning systems

__ Crash preparation systems

__ Navigation systems

__ Pedestrian detection

__ Phone tethering

__ Automatic brake drying

__ ________________________________

__ ________________________________

__ ________________________________

__ ________________________________

__ ________________________________

__ ________________________________

Long Form

Industry Input to the MIT AgeLab AAA-FTS Technology Rating Project – Technology Specific Information

Introduction to This Form - The Massachusetts Institute of Technology AgeLab has been tasked by the AAA Foundation for Traffic Safety to develop a comprehensive rating system for new in-vehicle technologies. AAA-FTS initiated this undertaking after observing that many consumers have a poor understanding of how new in-vehicle technologies work, and what benefits they may offer. The present focus of the system is on safety relevant technologies. A primary goal of this project is educational, to guide consumers towards more confident and strategic purchasing decisions that will enhance automotive safety by better educating them about new in-vehicle technologies that are most relevant and beneficial to their individual driving needs.

This form is intended to gather input from the automotive industry and traffic safety community that can be used to better develop these educational materials and to contribute to the identification of data and data sources that may be useful in the objective assessment of current safety technologies. **The information gathered through these forms will be combined with input from other industry sources to inform the overall project. Information and perspectives provided here will not be publically attributed to a specific manufacturer or source. Thus, while we view these forms as providing well-informed input from industry experts, they are not viewed as necessarily representing formal policy statements by a given manufacturer, industry representative, or other organization. These forms deliberately do not have a formal space for identification of the person or persons completing the form and responses can be treated entirely as confidential.** While the ability to follow-up with questions should they arise would be useful, identifying information to allow follow-up may be supplied at your own discretion. In addition, we anticipate that various stakeholders within an organization may have different perspectives. Where possible, submission of forms from multiple stakeholders within an organization is encouraged.

While this form may appear somewhat long at first, we recognize that many of the sections will not apply to all technologies and can simply be rated as "N/A". Similarly, if copies of these forms are provided to more than one individual or group within an organization, there may be a number of sections for which you have no background or specific expertise to comment upon. Please feel free to leave such sections blank.

If you cite technical reports, publications, or other supporting material that is technically public but may be hard to locate, please include them as attachments. If there are confidential reports you would like to supply, please reach out to us so that appropriate safeguards can be put in place before providing us with access to the material.

Finally, due to the aggressive schedule for this project, we would appreciate it if you can return individual forms as they are completed rather than waiting until all technologies have been reviewed or all stakeholders have responded. **Returning Forms** – Please return by e-mail to Dr. Bryan Reimer at: reimer@mit.edu

We would be glad to arrange a phone conversation to answer any questions or to discuss any topic areas in more detail. We can be contacted as below:

Bryan Reimer	617-452-2177	reimer@mit.edu
Bruce Mehler	617-253-3534	bmehler@mit.edu

For the Consumer

This section is intended to collect a consumer oriented, moderately detailed, mid-level overview of the technology. The subsection headings should be used as a guide for information we wish to communicate to the consumer. However, we are very open to considering any information or perspective that the industry feels would be useful in educating the public. You are encouraged to add proposed sections or material as you see fit. Be as brief or as expansive as you wish. Where possible, please reference (by number, name, or other indicator) empirical research or other objective data source for all assertions and add the source to the Reference List at the end of this document.

Note: This form is being supplied as a Word document so you can type information directly into the document.

Why Would I Use This Technology?

What is the purpose and major benefits of the technology? When asserting a benefit, please reference an empirical research source and include in the Reference List.

Insert here…

What Do Drivers Think?

Consumer survey and opinion data about the public's perception and use of the technology. As above, please cite sources.

How Well Does It Work?

Please summarize salient points about the efficacy of the technology, and go into a little more detail than the first subsection about relevant research. May include more detailed estimates of safety impact. Always cite sources.

Who Benefits Most?

Please highlight any driver demographics (families with small children, teens, older drivers) or situations (urban, rural, highway, night driving, icy environments, etc.) for which the technology is particularly beneficial.

In What Situations Doesn't It Work?

Drivers sometimes assume that a technology will provide protection in situations where it is not designed to function. Use this subsection to help educate the public about misunderstandings about a technology including technical limitations, conditions where it is not active, etc.

Mobility Significance

Does the technology have particular benefits for older drivers or persons with limited mobility?

Not All Systems Are Alike

If there are relevant differences between different implementations of the general technology type that impact what conditions they work under, relative effectiveness, etc., they should be highlighted here. (Technical or research data should be cited where possible to increase likelihood of inclusion.)

Different Names, Same Idea

Is the general technology known under a variety of industry names? If so, what are some of the common ones? (You are welcome to include brand specific names.)

Other

Please elaborate on any broad topic areas that are not addressed by the headings above. Feel free to add any additional information that you feel should be included to help educate the public about this technology.

The Underlying Knowledge Base

The section is intended to identify the underlying knowledge base and research that can be used to evaluate the actual safety benefit of a given technology. The information and data sources identified here will be combined with material gathered from other sources. As part of the educational component of the project, we anticipate also using this input in the development of technology review summaries that will be made available to industry, researchers, consumer advocates, etc. Draft versions of these summaries will be made available to contributors for review and comment before formal release. As noted earlier, material will compiled and individual comments and observations will not be attributed to specific contributors.

The intent here is to identify findings and data sources that can contribute to an objective evaluation of the function and effectiveness of a technology. Where empirical research or other objective data exists that is relevant to a category, please insert a brief summary statement highlighting the major finding , reference the source (by number, name, or other indicator), and then add the source to the Reference List at the end of this document.

If you are aware of research in a given category, but feel that the findings and/or conclusions are questionable or incorrect for some reason, please note this. Identifying questionable or poor research can be as important as identifying good work.

If there are no data sources that can be cited for a given category, please indicate if this is because no research appears to have been done (No Research) or if the category is not applicable / relevant to this technology (N/A). We are fully aware that for many of the technologies this may be the case and understand that there may be little or no details entered for a number of sections.

In summary:

- Entries can be as brief or as extensive as desired.
- Key statistics or findings should be supported by a citation / source that you add to the Reference List.
- If a factor area listed is not applicable to the technology, please explain why.
- Please feel free to comment on research reports or data sets that you believe misrepresent a technology (either underrating or overrating).

I. Scenario Significance

What issue, problem or risk is this technology attempting to address, and what statistics are available to quantify the significance of the problem? Are there relevant data from NHTSA or other sources that can be cited on number of crashes, injuries, fatalities, property damage associated with the issue the technology is intended to mitigate or eliminate?

Insert here…

II. Benefits & Theoretical Efficacy

Under What Conditions Is It Intended to Work? How Successful Is It?

What conditions is this technology designed to work under and what data is available indicating how successful it is at providing a benefit under these conditions? Be sure to enumerate specific safety benefits and supporting data. (If there are different versions of the technology that have different effectiveness characteristics, specify as needed.)

Other Benefits

If there are non-safety related benefits associated with the technology, they can be detailed as well. These might be convenience or comforted related, or might offer other benefits such as improvements in fuel economy.

Technical Limitations

Are there any conditions under which the technology will not operate, performance may degrade, or actual failure may occur (i.e. weather, speed, tolerance boundaries, etc.)?

Limitation Mitigation

For any of the limitations mentioned above, have any solutions been developed to offset, reduce, or remove these limitations?

Implementation Differences

Are there major differences between implementations of this general class of "technology" that need to be considered in evaluating its effectiveness, understanding or using the technology?

III. Human Interaction with the Technology

We recognize that there may be limited information available on how drivers interact with a particular safety technology and how this may influence its effectiveness. Please circle one of the summary ratings after each topic heading to indicate the extent to which you feel this aspect of human interaction with the technology is important. We understand that there may be little or no detailed information available to fill-in a number of the subsections.

Summary Importance Ratings: After each subheading, there are the headings:

N/A	not applicable to this technology
NI	not important
MI	moderately important
VI	very important

User Involvement (Please circle one: N/A, NI, MI, VI)

Does a driver need to know any anything about the technology to benefit from it or does it work largely or fully automatically?

Activation State and Use (Please circle one: N/A, NI, MI, VI)

As implemented in your vehicles, does the technology default on or does the user have to engage it? Does it default to a last used mode on start-up? If the user has to turn the technology on (or has the ability to turn it off), is there any data on the percentage of drivers who actively use the technology?

Incorrect Assumptions by Users (Please circle one: N/A, NI, MI, VI)

Are there misassumptions that drivers sometimes make about the technology? Assumptions that it will do certain things it is not designed to do? (These are important to identify from an educational perspective – if drivers make uninformed assumptions about what a technology should do that are outside of its design specification, a fundamentally good technology may get an undeserved poor reputation.) Both industry experience and identified research citations are appreciated here.

Consumer Awareness & Trust (Please circle one: N/A, NI, MI, VI)

What data is available on consumers' awareness of the technology, how it works, when they should and should not depend on it? Are there any data on consumers' level of satisfaction and trust in the technology? Survey data is acceptable here if citations to the data sources can be provided.

Behavioral Adaptation (Please circle one: N/A, NI, MI, VI)

This refers to behavior changes resulting from the use of the technology that may impact its net safety gain. Is there any suggestion of, or data on, behavioral adaption occurring or not occurring with this technology?

Demand Associated with the Technology

A technology may offer potential benefits while also placing certain demands on the driver before they can derive that benefit. Engaging a system may involve a degree of mental, visual, manipulative, or auditory workload. Attending to a warning may similarly require some amount of attention and resource allocation. What data are available on the extent to which the technology places some level of demand on the driver in each of the following domains? If a domain is not relevant to the implementation of this technology, please explain why.

Visual (Please circle one: N/A, NI, MI, VI) -

Auditory (Please circle one: N/A, NI, MI, VI) **-**

Manipulative (motor) (Please circle one: N/A, NI, MI, VI) **-**

Tactile (vibratory sensation) (Please circle one: N/A, NI, MI, VI) **-**

Cognitive (mental workload) (Please circle one: N/A, NI, MI, VI) -

Other -

Distraction or Confusion (Please circle one: N/A, NI, MI, VI)

Providing drivers with increased information in the form of added displays, warnings, automated corrections, etc. offer potential benefits but may also introduce some degree of distraction or confusion. What data are available on the extent to which this technology does or does not result in a degree of distraction or confusion in some drivers? What data are available on the extent to which a net positive benefit results from the technology?

Ease of Learning (Please circle one: N/A, NI, MI, VI)

Some systems require little or no familiarity with the technology to derive benefit from them. Others have a steep learning curve for a user to become comfortable with them, but may become second nature once the user has developed a good mental model of how they work. To what extent does the user need to learn how to use the system to derive benefit? Is there any data on how long most users take to become comfortable with the technology? Is there any data on the percentage of users who actively use the technology?

Stress (Please circle one: N/A, NI, MI, VI)

If a technology makes a driver more comfortable / less stressed driving under certain conditions, this may support an indirect safety benefit by increasing the driver's spare capacity to attend to the primary driving task. Is there any evidence that the technology increases or decreases driver stress in any way?

Other / Broad Comments on the Driver's Relationship to the Technology

Please use this space to share any additional information on how drivers appear to interact with the technology or to make any general comments on this topic area.

IV. Who Benefits Most? (Expanded)

This section can be used to expand upon the "Who Benefits Most?" segment of the consumer information form. Please highlight any driver demographics (families with small children, teens, older drivers) or *situations (urban, rural, highway, night driving, icy environments, etc.) for which the technology is particularly beneficial. What data is available to support these views?*

V. Mobility Significance (Expanded)

This section can be used to expand upon the "Mobility Significance" segment of the consumer information form. Does the technology have particular benefits for older drivers or persons with limited mobility? What data is available to support these views?

VI. External Safety

This refers to a technology's impact on the safety of persons or property outside of the immediate driver's vehicle, such as pedestrians and other vehicles. For example, a low-speed collision detection system that automatically slows the vehicle may reduce the severity of pedestrian injuries, conferring a benefit outside of the immediate vehicle. Data on both positive and negative external safety considerations should be noted here.

VII. Other Considerations

Please use this area to comment on any other factors that you feel should be taken into account in evaluating the safety benefits of this technology? What data is available to evaluate this technology on any such factors?

Research Needs

In addition to the educational objectives of this project, another goal is to identify areas where additional research is needed in the assessment and optimization of current and emerging safety technologies. Work in these areas is carried out individually and jointly by partnerships involving industry, academia, foundations, and government. If you feel that there are research needs specific to this technology, or that apply more generally to a safety related issue, that would be useful to promote as a research agenda, please comment below.

References

Please list here all empirical research or other objective data sources that are used to back-up statements made about a technology. Technical specification documents and industry reports, if they can be made available for review, can be cited for this purpose. If you cite technical reports, publications, or other supporting material that is technically public but may be hard to locate, please consider including them as attachments.

Please feel free to cite sources in any format that is convenient, from web links to formal academic citations.

Short Form

Industry Input to the MIT AgeLab AAA-FTS Technology Rating Project: Technology Specific Information

This form is a short version of the full *Technology Information Form* that was developed to allow the industry to provide detailed input to the MIT AgeLab AAA-FTS Technology Rating Project. The primary intent of this information support request is to ensure that the industry has an ample opportunity to:

1. help identify information that can be drawn upon to establish the extent to which various new and emerging technologies provide a safety benefit to the consumer, and
2. contribute to the framing of educational materials that can help better inform the public about why they might wish to consider purchasing vehicles with various technologies given their particular needs and individual driving considerations.

While a number of industry sources have commented positively on the comprehensive nature of the full form, some have indicated that they would like to contribute but may not have the ability at this time to make full use of the complete form. We want to make it clear that contributing selective input to this process is more meaningful than not contributing at all. As noted in the information supplied with the full form, respondents should feel free to fill-in or skip sections based on availability of expertise, time, etc. Similarly, while commenting on all the technologies being included in the first round review provides a given contributor the widest opportunity to have their perspective represented, selecting to comment on a targeted subset of technologies of greatest interest to your group or company is certainly a reasonable option.

To encourage the widest possible input to this project, we are providing this short version of the information input form as an optional method of contributing. Please consider using this short form for commenting on technologies that you are not able to review in detail using the full form, or as an alternate approach overall to sharing what you consider to be important information or descriptive material.

We would be glad to arrange a phone conversation to answer any questions or to discuss any topic areas in more detail. We can be contacted as below and forms can be returned to Dr. Bryan Reimer at reimer@mit.edu.

Bryan Reimer	617-452-2177	reimer@mit.edu
Bruce Mehler	617-253-3534	bmehler@mit.edu

Technology

Please indicate below what technology you are commenting on in this form.

__ Lane Departure Warning

__ Back-up Cameras

__ Forward Collision Warning

__ Forward Collision Mitigation

__ Adaptive / Smart Cruise Control (headway management)

__ Adaptive Headlamps

__ Electronic Stability Control

__ Other: ________________________

Responses can be as brief or as extensive as you would like to contribute to this project. A partially completed form has more impact than no response.

Please provide a short description that highlights for the consumer what this technology is intended to do and why they would benefit from having and using the technology. *(What would you like to say to the consumer about this technology through the educational medium of this project?)*

Who is likely to benefit the most from this technology? *(Are there any driver demographics (families with small children, teens, older drivers) or situations (urban, rural, highway, night driving, icy environments, etc.) for which the technology is particularly beneficial?)*

Is there anything the consumer should understand about how this technology works and what it does and does not do? (*Drivers sometimes assume that a technology will provide protection in situations where it is not designed to function. Use this section to help educate the public about misunderstandings about a technology including technical limitations, conditions where it is not active, etc.)*

What objective data is available to support the position that this technology provides an actual safety benefit? *(We are looking for input on both what is available in the public domain that can be reviewed as well internal data that might be shared for purposes of this project. What convinced you / your company that this technology was worth investing in? What has been learned over time that makes the case for a safety benefit?)*

Are there features of particular implementations of this class of technology that the public should be made aware? *(While this project is currently oriented toward educating the public about a technology as a class as opposed to rating individual implementations, there may be features you feel should be part of the discourse. If there are relevant differences between various implementations of the general technology type that impact what conditions they work under, relative effectiveness, etc., they should be highlighted here. Any mitigation solutions developed to deal with potential limitations can be highlighted. Technical or research data should be cited where possible to increase likelihood of inclusion.)*

Research Needs – Are there any gaps in our current understanding of these technologies where additional research would be useful? *(In addition to the educational objectives of this project, another goal is to identify areas where additional research is needed in the assessment and optimization of current and emerging safety technologies. Work in these areas is carried out individually and jointly through partnerships involving industry, academia, foundations, and government. If you feel that there are research needs specific to this technology, or that apply more generally to a safety related issue, that would be useful to promote as a research agenda, please comment below. For example, are there enhancements in sensor technology that are needed? Are there human factors considerations related to how drivers perceive and interact with this technology that would be useful to understand more fully?)*

Ergonomics

Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/terg20

Multi-Modal Assessment of On-Road Demand of Voice and Manual Phone Calling and Voice Navigation Entry across Two Embedded Vehicle Systems

Bruce Mehler[a], David Kidd[b], Bryan Reimer[a], Ian Reagan[b], Jonathan Dobres[a] & Anne McCartt[b]

[a] MIT AgeLab & New England University Transportation Center, 77 Massachusetts Avenue, E40-291, Cambridge, MA 02139, USA, Tel: (617) 253-3534, Fax: (617) 258-7570

[b] Insurance Institute for Highway Safety, 1005 N. Glebe Road, Arlington, VA 22201, USA

Accepted author version posted online: 13 Aug 2015.

To cite this article: Bruce Mehler, David Kidd, Bryan Reimer, Ian Reagan, Jonathan Dobres & Anne McCartt (2015): Multi-Modal Assessment of On-Road Demand of Voice and Manual Phone Calling and Voice Navigation Entry across Two Embedded Vehicle Systems, Ergonomics, DOI: 10.1080/00140139.2015.1081412

To link to this article: http://dx.doi.org/10.1080/00140139.2015.1081412

Embedded Vehicle Voice Systems (2015-06-16)

Publisher: Taylor & Francis
Journal: *Ergonomics*
DOI: http://dx.doi.org/10.1080/00140139.2015.1081412

Multi-Modal Assessment of On-Road Demand of Voice and Manual Phone Calling and Voice Navigation Entry across Two Embedded Vehicle Systems

Bruce Mehler*[1], David Kidd[2], Bryan Reimer[1], Ian Reagan[2], Jonathan Dobres[1], & Anne McCartt[2]

[1]*MIT AgeLab & New England University Transportation Center*

77 Massachusetts Avenue, E40-291, Cambridge, MA 02139, USA

[2]*Insurance Institute for Highway Safety*

1005 N. Glebe Road, Arlington, VA 22201, USA

*Corresponding author. Email: bmehler@mit.edu
phone: (617) 253-3534 fax: (617) 258-7570

Funding: This work was supported by the Insurance Institute for Highway Safety and a grant from United States Department of Transportation's Region One University Transportation Center at MIT.

Disclosure Statement: The authors have no financial interest or benefit arising from the direct application of this research.

Multi-Modal Assessment of On-Road Demand of Voice and Manual Phone Calling and Voice Navigation Entry across Two Embedded Vehicle Systems

One purpose of integrating voice interfaces into embedded vehicle systems is to reduce drivers' visual and manual distractions with "infotainment" technologies. However, there is scant research on actual benefits in production vehicles or how different interface designs affect attentional demands. Driving performance, visual engagement, and indices of workload (heart rate, skin conductance, subjective ratings) were assessed in 80 drivers randomly assigned to drive a 2013 Chevrolet Equinox or Volvo XC60. The Chevrolet MyLink system allowed completing tasks with one voice command, while the Volvo Sensus required multiple commands to navigate the menu structure. When calling a phone contact, both voice systems reduced visual demand relative to the visual-manual interfaces, with reductions for drivers in the Equinox being greater. The Equinox "one-shot" voice-command showed advantages during contact calling but had significantly higher error rates than Sensus during destination address entry. For both secondary tasks, neither voice interface entirely eliminated visual demand.

Practitioner summary: The findings reinforce the observation that most, if not all, automotive auditory-vocal interfaces are multi-modal interfaces in which the full range of potential demands (auditory, vocal, visual, manipulative, cognitive, tactile, etc.) need to be considered in developing optimal implementations and evaluating drivers' interaction with the systems.

Keywords: Voice interface; visual demand; distraction; workload; human machine interface

Social Media (140 characters): In-vehicle voice-interfaces can reduce visual demand but do not eliminate it and all types of demand need to be taken into account in a comprehensive evaluation.

1. Introduction

Manufacturers are equipping vehicles with embedded systems that allow occupants to interact with entertainment, communication, and driver support systems built into the vehicle (e.g., radio, navigation) and connected portable devices (e.g., cell phones, MP3 players). Increasingly, embedded systems allow occupants to interact with different functions or devices with voice commands in addition to traditional visual-manual interactions using buttons, knobs, or a touchscreen. A perceived advantage of voice inputs compared with manual inputs is that they eliminate or reduce the competition for visual and manual resources between a secondary activity and the primary task of driving. Therefore, voice interfaces have been widely considered as an appealing approach for giving drivers access to a range of entertainment and connectivity options while minimizing the potential impact on driving performance and safety. At the same time, there remains a concern that performing any secondary task can increase crash risk, and some caution that adding even easy-to-use interfaces may raise the total amount of attention drivers give to non-driving tasks. Regardless, a deeper understanding of the various demands originating from drivers' interactions with voice interfaces is needed to more objectively optimize tasks in which drivers engage and to identify tasks that can lead to an inappropriate level of disruption in driving.

The apparent benefits of using voice inputs to interact with a device while driving compared with manual inputs are well documented in experimental research using various simulated driving performance measures. For instance, the standard deviations of lane position and reaction time are not as degraded relative to baseline driving, and may be less than baseline driving, when drivers perform a secondary auditory-vocal task versus a secondary visual-manual task (e.g., Haigney, Taylor, & Westerman, 2000; Maciej & Vollrath, 2009; Ranney, Harbluk, & Noy, 2005; Tsimhoni, Smith, & Green, 2004). Schreiner, Blanco, and Hankey (2004) have shown parallel findings with a simulated voice system on a closed roadway. Drivers also look away from the roadway less often and for less time during voice interactions (Chiang, Brooks, & Weir, 2005; Mehler, Reimer, Dobres, McAnulty, Mehler, Munger et al., 2014; Owens, McLaughlin, & Sudweeks, 2010; Reimer, Mehler, Dobres, & Coughlin, 2013; Reimer, Mehler, Dobres, McAnulty, Mehler, Munger et al., 2014; Schreiner, Blanco, & Hankey, 2004; Shutko, Mayer, Laansoo, & Tijerina, 2009). Attempts to use voice interactions as a means of reducing the amount of time that a driver's eyes are directed away from the road is easily understood. In studies of small samples of drivers who were continuously monitored over an extended period of time, the risk of a crash, near-crash, or other type of safety conflict increased the longer the driver's eyes were off the road (e.g., Klauer, Dingus, Neale, Sudweeks, & Ramsey, 2006). Thus, systems that allow drivers to look away from the roadway less often may degrade safety less. However, experimental research also indicates that some voice interactions affect driving performance in ways that could increase crash risk. Speech generation, speech comprehension, and even simple cognitive tasks can affect simulated driving performance (e.g., speed variability, lane maintenance) and mental workload, especially when the information is complex or poorly implemented within the vehicle (e.g., Blanco, Biever, Gallagher, & Dingus, 2006; Kubose, Bock, Dell, Garnsey, Kramer, & Mayhugh, 2006; Lee, Caven, Haake, & Brown, 2001; Strayer, Cooper, Turrill, Coleman, Medeiros-Ward, & Biondi, 2013).

1.1 On-road research with production voice systems compared with visual-manual interaction

Although the potential benefits of voice-based interactions compared with visual-manual interactions are well documented in simulated or prototype implementations (see previous section and reviews by Barón & Green, 2006; Lo & Green, 2013; Reimer et al., 2013), only a few studies have examined if these benefits exist with production-level embedded systems (e.g., Graham & Carter, 2000; Harbluk, Burns, Lochner, & Trbovich, 2007; Owens et al., 2011; Shutko et al., 2009), and even fewer have examined the use of production systems on actual roadways. Chiang, Brooks, and Weir (2005) conducted two studies where drivers entered a destination into a navigation system using a point-of-interest selection, the destination's phone number, or the street address with similar built-in navigation systems in a 2004 Accord and 2005 Acura RL. Participants completed the destination entry tasks using voice or manual inputs while driving on city streets and a freeway. Drivers spent a smaller percentage of time looking at the navigation system interface, had less variability in lane keeping performance, and reported lower subjective ratings of mental workload using voice inputs compared with manual inputs. In another on-road study, Owens, McLaughlin, and Sudweeks (2010) reported that using the embedded voice interface of the Ford SYNC® system to complete several infotainment tasks lowered drivers' visual demand, steering variability, and subjective mental workload relative to using a portable hand-held cell phone. Owens et al. (2010) did not examine performance during manual interactions with the vehicle's embedded system or voice interactions with the cell phone.

In a series of studies, Reimer, Mehler, and colleagues (Reimer & Mehler, 2013; Mehler et al., 2014; Reimer, Mehler, Dobres, McAnulty, Mehler, Munger & Rumpold, 2014) also examined Ford SYNC® and an embedded navigation system in a 2010 Lincoln MKS. While driving on an interstate highway, drivers used voice commands in a set of navigation, phone calling, and entertainment tasks. As comparison points, drivers also manually changed radio stations using preset buttons, engaged in a more intensive visual-manual reference task of locating specific radio stations that required multiple button presses and rotation of the tuning knob, and completed several levels of a working memory cognitive reference task. In line with the studies cited above that showed some advantages for voice-command systems, drivers looked away from the roadway less when using the voice interface to select a radio station than when using the multi-step manual radio tuning interface. On the other hand, compared with multi-step manual radio tuning, drivers looked away from the roadway for substantially longer periods of time when using voice inputs to enter a destination address into the navigation system or attempted to select a song that did not exist in the entertainment system. Reimer and Mehler (2013) noted that the user interface for voice input presented information on the center console display (e.g., voice-command options, street or city name selection options), which inherently provides a reason to look at the screen.

Together, the findings from these on-road studies suggest that, compared with manual interaction, voice interaction with embedded and portable systems can reduce visual demand as intended, but do not necessarily eliminate it. Some voice interactions appear to result in moderate to large visual engagement when considered in terms of metrics such as glance time to device or total eyes-off-road time that have been used or proposed for evaluating visual-manual interfaces (Driver Focus-Telematics Working Group, 2006; National Highway Traffic Safety Administration (NHTSA), 2013).

1.2 Variation across system implementations

However, the designs of embedded systems vary, and some interface designs may be more effective at minimizing visual demand than others. Reagan and Kidd (2013) used hierarchical task analysis to count the steps required to dial a 10-digit phone number, dial a contact in a cell phone contact book, and tune to a radio station to identify differences in the manual and voice interfaces of four embedded systems in 2013 model year vehicles. Two distinct design approaches for voice interactions emerged from this static evaluation. The first was a menu-based approach where tasks were completed by using contextual voice commands to progress through a series of menus and submenus, often mimicking the sequence of manual inputs required to complete the task. The second was a “one-shot” approach where a single compound command was used to execute most or all of the task in a single step.

As initially observed in Reagan and Kidd (2013), the differences in these two approaches were most apparent between the Chevrolet MyLink and Volvo Sensus systems. For example, calling a contact in the phone book could be performed using a single voice command with the Chevrolet MyLink (e.g., the driver saying “Call home,” which resulted in the system response, “calling home on cell” and initiation of the phone call). The same task required four separate voice commands with the Volvo Sensus as the user moved through different menus and verified previous commands (e.g., to call “home” a driver said “Phone, call contact;” waited for the system prompt “name please;” said the contact name “home;” waited for the system prompt, “please say a line number;” said “one;” waited for the system prompt “dial home mobile – confirm: yes or no;” and then said “yes,” after which Sensus made the call). Calling a contact with the Volvo Sensus took more steps when using voice inputs than when using manual inputs. Furthermore, many of the system prompts asked the driver to look at the center stack display to choose among options for the contact (e.g., home, cell, work) or to confirm input, and the prompt asking for a line number always occurred whether the contact (“home’ in the example here) had one or multiple line numbers. This integration of visual information to support the voice interface was similar to that noted by Reimer and Mehler (2013) in their initial evaluation of the Ford SYNC® system. Providing visual information to support voice input may help alleviate the cognitive demand associated with memorizing and recalling voice commands, or an auditory list of options. However, it is unclear whether and to what extent safety or other trade-offs are associated with reducing cognitive demand at the expense of increased visual demand.

While calling a contact in the phone book with Chevrolet MyLink required fewer voice commands than Volvo Sensus, MyLink required a deeper understanding of system operation in that it did not provide as much prompting, visual support, or confirmation, which could potentially result in more calling errors. The cognitive demand associated with recalling complex voice commands might negate the benefits associated with reducing overall task duration and the potential for visual engagement. Previous research has shown that cognitive demand from a secondary task can interfere with visual information processing (Just, Keller, & Cynkar, 2008; Strayer, Drews, & Johnston, 2003) and constrict visual scanning patterns (Recarte & Nunes, 2000; Reimer, Mehler, Wang, & Coughlin, 2012).

A recent study by Garay-Vega, Pradhan, Weinberg et al. (2010) found that the differences between a menu-based voice interface and one-shot voice interface are not negligible. Drivers completed a music retrieval task using an iPod™, a multiple turn voice interface (i.e., menu-based voice interface), and a single turn voice interface (i.e., one-shot voice interface) during simulated driving. The task took longer to complete using the multiple turn voice interface. Furthermore, only the single turn voice interface reduced the average time that drivers had their eyes off the road compared with the iPod™. The multiple turn voice interface was also perceived to be more demanding than the single turn interface.

In sum, naturalistic driving research indicates increased risk of safety-critical events from the visual and manual demands of in-vehicle secondary tasks (Fitch, Soccolich, Guo, McClafferty, Fang, Olson et al., 2013; Klauer et al. 2006; Victor, Bärgman, Boda, Dozza, Engström, Flannagan et al., 2014). Research also indicates that voice interfaces reduce workload and visual attentional demand relative to visual-manual interfaces (e.g., Chiang et al. 2005; Haigney et al., 2000; Mehler et al., 2014; Owens et al., 2010; Reimer et al., 2013; Shutko et al., 2009). However, recent research indicates that voice-based interactions may introduce noticeable visual demand (e.g., Mehler et al., 2014; Reimer et al., 2013) and that some voiced interface designs can increase perceived workload and visual demand when driving in a simulator relative to others (e.g., Garay-Vega et al., 2010). These findings support concerns that distraction can remain (depending upon implementation), despite the use of voice-based systems and led to the task analysis by Reagan and Kidd (2013), described above, and to the study reported here.

1.3 Objectives & approach

A primary objective of the current study was to compare the relative demands of production implementations of primarily visual-manual vs. voice-involved human machine interfaces intended to allow completion of the same end-goal task while driving by considering the effects on driving performance, visual demand, and indices of mental workload (heart rate, skin conductance, and subjective ratings). Of equal interest was an exploration of the significance of differing design approaches to voice-based systems (e.g., a one-shot vs. multi-step entry). A 2013 Chevrolet Equinox equipped with the MyLink system and a 2013 Volvo XC60 equipped with Sensus served as test case exemplars of these two system designs in the research reported here.

Volunteer drivers drove either the Chevrolet or Volvo on a highway while initiating calls through a phone contact list using voice and manual inputs and entering addresses into the navigation system using voice input with the vehicle's embedded system and a mounted smartphone. In the case of phone calling, using voice inputs of the embedded systems was expected to degrade driving performance less, reduce visual demand, and lower workload levels compared with performing these tasks manually. Based on the task analysis by Reagan and Kidd (2013), the relative benefits of using voice input compared with manual input were expected to be greater for drivers using the Chevrolet MyLink. However, the absence of verification steps with MyLink was expected to increase the number of errors using voice inputs for complex tasks such as address entry.

2. Methods

2.1 Participants

Participants were identified primarily using online and newspaper advertisements in the greater Boston area. Recruitment was directed at obtaining a sample of relatively healthy and experienced drivers. Participants were required to be between the ages of 20 and 69, have been licensed for a minimum of 3 years, and self-report driving at least 3 times a week and being in relatively good health for their age. Also based on self-report, individuals were excluded if they were a driver in a police-reported crash in the past year, were positive for any of a range of serious medical conditions (e.g., a major illness resulting in hospitalization in the past 6 months, a diagnosis of Parkinson's disease, a history of stroke) or were taking medications that might impair their ability to drive safely under the study conditions (e.g., anti-convulsants, anti-psychotics, medications causing drowsiness).

Recruitment continued until a sample of 80 participants with usable driving performance, glance and physiological data, and equally balanced across the two vehicles by gender and age was obtained. The target age distribution was in general conformity with the recommendations for device assessment in NHTSA's (2013) *Visual-Manual Driver Distraction Guidelines for In-Vehicle Electronic Devices*, which call for an equal number of participants across four age groups (18-24, 25-39, 40-54, 55 and older). In line with recruitment goals, participant age did not vary significantly by gender or vehicle (both $F(1,79) = .949$)(see

Table **1**); the sample ranged in age from 20 to 66 years. Recruitment procedures and the overall experimental protocol were approved by MIT's institutional review board, and compensation of $75 was provided.

Table 1. Mean age (and SD) of participants by age group, gender, and vehicle.

Age Group	Chevrolet (n=40) Female (n=20)	Chevrolet (n=40) Male (n=20)	Volvo (n=40) Female (n=20)	Volvo (n=40) Male (n=20)	Combined n=80
20-24 (n=20)	21.4 (0.9)	22.4 (1.8)	23.2 (0.8)	21.2 (1.3)	22.1 (1.4)
25-39 (n=20)	33.0 (3.4)	31.2 (4.9)	28.8 (3.2)	28.9 (4.0)	30.5 (4.3)
40-54 (n=20)	45.4 (4.0)	47.6 (3.9)	48.6 (5.0)	49.8 (3.7)	47.9 (4.2)
55-69 (n=20)	62.4 (2.7)	59.0 (2.6)	59.8 (4.0)	62.4 (4.3)	60.9 (3.6)
Combined	40.6 (15.8)	40.1 (14.9)	40.2 (16.0)	40.6 (17.1)	40.3 (15.6)

2.2 Apparatus

Participants drove one of two standard production vehicles, a 2013 Chevrolet Equinox equipped with the MyLink system and a 2013 Volvo XC60 equipped with the Sensus system. Both vehicles were equipped with forward collision warning and lane departure warning safety systems. Phone connectivity was supported by pairing a Samsung Galaxy S4 smartphone (model SCH-1545) to each vehicle's embedded system via the vehicle's Bluetooth wireless interface. Both vehicles were instrumented with a custom data acquisition system for time synchronized recording of vehicle information from the controller area network (CAN) bus, a Garmin 18X Global Positioning System (GPS) unit, a MEDAC System/3™ physiological monitoring unit to provide EKG and skin conductance level (SCL) signals, video cameras, and a wide-area microphone to capture driver speech and audio from the vehicle's speech system. The five video cameras provided views intended to capture the driver's face for primary glance behavior analysis, the driver's interactions with the vehicle's steering wheel and center console, the forward roadway (narrow and wide-angle images), and a rear roadway view. Data were captured at 10 Hz for the CAN bus and GPS, 30 Hz for the face and narrow forward roadway cameras, 15 Hz for the remaining cameras, and 250 Hz for the physiological signals to support EKG feature extraction for heart beat interval detection.

For EKG recordings, the skin was cleaned with isopropyl alcohol and standard pre-gelled silver/silver chloride disposable electrodes (Vermed A10005, 7% chloride wet gel) applied using a modified lead II configuration that placed the negative lead just under the right clavicle, the ground just under the left clavicle, and the positive lead over the lowest left rib. Skin conductance was measured utilizing a constant current configuration and non-polarizing, low impedance gold plated electrodes that allowed electrodermal recording without the use of conductive gel. Sensors were attached with medical grade paper tape on the underside of the outer segments of the middle fingers of the left hand to leave the right hand free for engaging the push-to-talk button on the steering wheel of each vehicle and controls on the instrument cluster. The thin surface design of the electrodermal sensors minimized interference with a natural grip of the steering wheel associated with the use of more traditional cup style electrodes.

2.3 Secondary Tasks

2.3.1 Calling a phone contact

A phone list of 108 contacts was used for all phone calling tasks. Characteristics of how each system organized information were taken into consideration so that neither system was disadvantaged. The list was ordered by first name and entries started with A and ranged through R so that all target selections could be reached through a comparable number of manual actions in each system. There were 18 names in each of 6 alphanumeric ranges corresponding to the bin organization used in the Chevrolet MyLink manual interface (ABC, DEF, GHI, JKL, MNO, PQRS, TUV, and WXYZ).

Calling a phone contact was presented as a sequence of two 'easy' and two 'hard' tasks. The easy tasks were calling a contact with only one phone number entry (Mary Sanders and Carol Harris). The hard tasks were calling a contact with two phone numbers (e.g., home and mobile). For these contacts (Pat Griffin on mobile and Frank Scott at work), the target phone was never the first listing so that simply requesting the contact name alone would not dial the correct number. The form of the easy task prompt was, 'Your task is to call Mary Sanders. Begin.' The form of the hard task prompt was, 'Your task is to call Frank Scott at work. Begin.' The contacts were the same across the manual and voice interface interactions so that any aspect/characteristic of a particular contact name that might influence the relative difficulty was constant (e.g., alphabetic location). As previously noted, a detailed description of the operations and resources required to dial a contact using the MyLink and Sensus systems is provided in Reagan and Kidd (2013). The key elements of each approach as they relate to the tasks used in this study follow.

Calling a contact using the MyLink visual-manual interface had the most discrete steps and began by locating and selecting the phone subsystem, followed by selecting the alphanumeric bin (e.g., ABC, DEF) containing the target contact. For contacts with a single phone number (easy case), the contact name was then selected from the list. In the case of two numbers for a contact name (hard case), both numbers appeared sequentially in the same list (i.e., Frank Scott work, Frank Scott home) and the target option was selected. Calling a contact using the Sensus visual-manual interface required the user to select the phone subsystem and then scroll through the upper level of the contact list to the appropriate contact name using a rotary knob on the center console. In the case of contacts with a single phone number (easy case), pressing an "OK" button initiated the call. For contacts with multiple numbers (hard case), pressing the button brought up a submenu listing the phone numbers for that contact. The rotary dial was again used to locate the desired selection and the "OK" button was pressed.

In contrast to manual calling, the MyLink voice interface required few steps. After pressing the push-to-talk button on the steering wheel, the driver could initiate both the easy and hard tasks in a single command string (i.e., 'call Mary Sanders,' 'call Pat Griffin on mobile'). No confirmation step was required if the system had confidence in the identification of the selection. This kept the interaction brief but meant the driver had to interrupt call initiation if a recognition error occurred. The voice implementation in the Sensus system more closely mirrored the multi-level menu structure used in the manual interface and asked for confirmation of selections. In specific, after pressing the push-to-talk button on the steering wheel, the driver could issue the compound command ‘Phone call contact’ to access the phone list and then say ‘Mary Sanders.’ The entry list would then appear on the display screen, and the driver was asked to say a line number and then asked to confirm the selection. In the case of multiple phone numbers for the contact, a second level menu would appear showing the options. The driver selected from this listing verbally and then confirmed the selection.

Each phone number associated with a target contact connected with a voicemail recording that confirmed the contact identity and stated that the phone call could now be disconnected. If the target contact was not reached, the call connected to a voicemail indicating that the MIT AgeLab had been reached and the phone call could now be disconnected. This provided auditory confirmation to the participant and research associate as to whether the target contact had been correctly selected or not.

2.3.2 Entering an address into the navigation system

Participants were asked to enter three addresses into each navigation system: 1) 177 Massachusetts Avenue, Cambridge, Massachusetts; 2) 293 Beacon Street, Boston, Massachusetts; and 3) their home address. The form of the prompt was, ‘Your task is to enter the destination address: 177 Massachusetts Avenue, Cambridge, Massachusetts. Begin.’ The first two addresses also were printed in large black text on a white card attached to the center of the steering wheel to minimize any cognitive load of needing to memorize and hold the address in memory during the duration of the interaction with the navigation system. This card was in place throughout the drive so that participants were exposed to the addresses for a minimum of 40 minutes prior to being asked to enter them into the system.

Address entry with MyLink was initiated by pressing the “push-to-talk” button and saying the command “navigation.” After prompting the driver for a navigation command, the system was flexible in terms of command syntax, accepting variations including “destination address,” “enter address,” and simply “address.” The full address could then be entered as a single verbal string in the form “177 Massachusetts Avenue, Cambridge, Massachusetts.” If the system was able to parse the string into component parts that was interpreted as a unique address at a high confidence level, there was no confirmation step and navigation instructions were initiated. If multiple potential targets were identified, they were presented auditorially and visually to be selected by verbalizing an option number.

Address entry using Sensus supported the compound but specific command, “navigate go to address” to select address entry. In contrast to the single string “one shot” approach of MyLink, Sensus prompted the user for the component parts of the address in steps, i.e., city name, street name, and street number were entered separately. Recovery from a user error or system misidentification at each step required little familiarity with the system as the prompt for each step offered the option of returning to a previous step, e.g., “please say the house number in single digits or say correction.” If the street number was correctly identified, the driver was prompted to say “finish.” An additional confirmation step prompted the driver to say “enter destination” to proceed with initiating navigation. If the system identified multiple potential targets, a list of options was shown on the center stack display screen and the system prompted the driver to “say a line number or say not on list.”

2.3.3 Instructions on task prioritization

Participants were instructed several times (in the written consent form, by recorded instructions, and through direct prompting by the research associate in the vehicle) that priority should be given to safe driving. Recorded instructions presented just prior to starting the drive stated the following: “*When you reach I-495 and have had a few minutes of driving on that highway, short recorded prompts will tell you what task we would like you to consider trying. When you hear these prompts, please do not start a task until you hear the word ‘begin.’ While we would like you to consider doing each task, you should always give priority to safe driving. If you feel for any reason that a task will interfere with your ability to drive safely, delay starting the task until you feel it is safe to do so or skip the task entirely if you feel that is the best thing to do. Your safety, and the safety of other people around you, is the highest priority. Please also be aware that you will be responsible for paying for any citations that you might be issued for traffic regulation violations. If at any time you feel uncomfortable driving the vehicle or in your ability to drive safely, please let the research associate know how you are feeling and they will confer with you about pulling off the roadway at the nearest safe location.*”

2.4 Experimental design

Gender- and age-balanced samples were distributed across the Chevrolet and Volvo vehicles (see

Table 1). As represented schematically in Figure 1 and further detailed in section 2.6 on procedure, participants were presented with the phone calling tasks to be undertaken using a voice-based interface and a visual-manual interface, and presented with addresses to enter into a voice-based navigation interface. While the present report focuses on the use of embedded vehicle systems to engage in these tasks, participants were also presented with the same tasks to be accomplished using a smartphone. Within each vehicle group, random assignment was made to either an “embedded vehicle system” or a “smartphone” first condition. Within each condition, random assignment determined whether voice-based or manual phone calling was presented first. Consequently, any advantage of being presented with the same contact to dial a second time was balanced across the modalities. The address entry tasks were always presented between the two methods of phone calling.

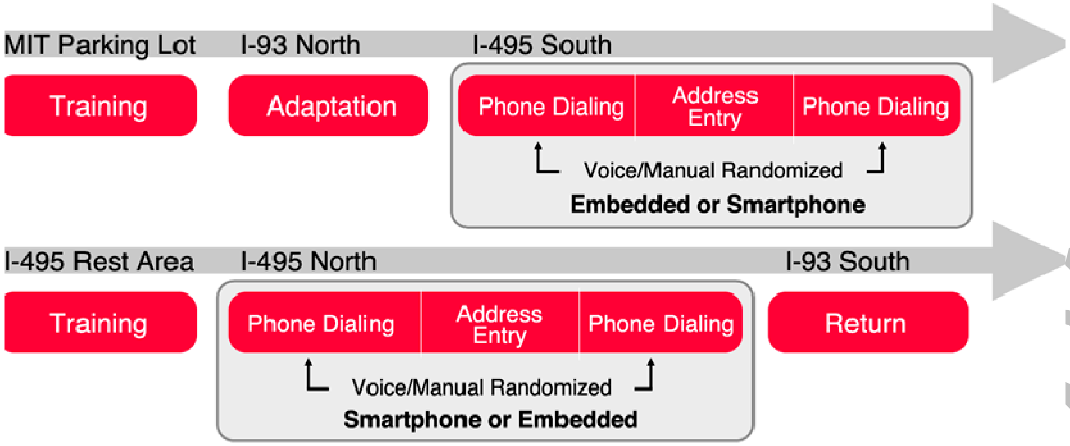


Figure 1: Schematic representation of the experimental design. Half of the participants interacted with the embedded vehicle system (Chevrolet MyLink or Volvo Sensus) during the first half of the drive and half during the second.

In sum, across 6 distinct task periods, each participant was presented a total of 22 secondary tasks, 11 during the southbound segment using either the embedded or smartphone device (4 manual phone calling trials, 3 address entry trials, and 4 voice calling trials) and then repeated the same 11 tasks during the northbound segment using the alternate device. As noted above, the present paper focuses on the participants' interactions with the embedded systems. Findings considering interactions with the Smartphone are presented in a companion report (Reimer, Mehler, Reagan, Kidd, & Dobres, 2015); however, a description of when participants were trained and assessed on both the smartphone and embedded systems is included in the following summary of the study procedures to provide a complete overview of the larger study.

2.5 Procedure

Participants reviewed and signed an informed consent, and a structured face-to-face interview confirmed eligibility that was initially established via phone or on-line screening. A questionnaire covering demographic information, attitudes toward driving, and technology experience was completed, an explanation of the workload rating scale provided, and physiological sensors were attached.

Participants were escorted to the research vehicle, instructed on how to adjust the seat and mirrors, asked to back-up the vehicle and then return to the parking space to obtain an initial feel for the vehicle, and encouraged to make additional adjustments for comfort and visibility as desired. Participants were trained in the parking lot in the use of the system (embedded or smartphone) to which they were assigned for the first half of the drive. Training began with manual phone calling, followed by voice phone calling, and then by voice destination address entry. Participants were trained to use the relevant rotary knobs on each vehicle's center stack. The Chevrolet had a touch screen interface that could be used to complete selection tasks as an alternative to pressing the rotary knob; similarly, the Volvo had a thumb wheel on the steering wheel that could be used as an alternative method for scrolling through lists and making selections. Although they were not trained to use these alternative methods, participants were allowed to use the method if they discovered and preferred its use over the rotary center stack knob. For the embedded vehicle systems, following the approach taken in Reimer et al. (2013) and Mehler et al. (2014), the default factory settings for the vehicle voice interfaces were used. Moreover, participants were given guidance on the use of short-cut command options to reduce the number of steps required to complete tasks. As an example, to use the voice interface in the Sensus system, calls could be placed by saying the upper level command 'Phone", waiting for a response, and then saying 'Call Contact". During training, participants where told "Calls can be placed by speaking the command 'Phone Call Contact'; you can also use the shorter command, 'Call Contact'." The remainder of the training interaction then focused on the shorter version.

As just described, training focused on providing participants with guidance on an efficient method of completing the tasks, while not constraining them to a fixed set of steps (beyond using the voice or manual interface at specified times) if the interface allowed multiple ways of accomplishing the task. Allowing participants to select an alternate method (e.g., touching a selection on-screen rather than pressing the rotary control) if it felt more comfortable to them was seen as an approach that made task performance more naturalistic. This approach does diverge from an assessment that is specifically aimed at quantifying the demand associated with an exact sequence of steps as might be done in testing whether a specific method of completing a task meets the NHTSA (2013) voluntary guidelines.

Participants with the Volvo Sensus system were taken through the voice calibration procedure, which is intended to tune the voice recognition system to participants' pronunciation based on a set of command relevant words; the Chevrolet MyLink system did not offer this feature. Orientation and training consisted of recorded instructions to provide consistency, supplemented with guidance by a research associate to clarify details and answer questions as needed. Participants were encouraged to repeat tasks until they felt comfortable to proceed. The orientation/training period typically ranged between 15 and 30 minutes, with a mean of approximately 20 minutes.

Then participants drove the vehicle on actual roadways in and around the greater Boston area. A driving adaptation period of approximately 30 minutes took place prior to the start of the formal assessment. The route consisted of approximately 10 minutes of urban driving from MIT to interstate highway I-93 and approximately 20 minutes north on I-93 to I-495. For the portions used in this study, I-495 is a divided interstate that is largely surrounded by forest with three traffic lanes in each direction with lane widths of 15 feet (3.62 m). The posted speed limit is 65 mph (104.6 kph).

Presentation of the secondary tasks with the first assigned system interface (smartphone or embedded system) occurred while driving south on I-495 (see Figure 1). At the end of this southbound segment, a break was taken at a highway rest stop where participants completed workload and other ratings for the tasks just completed. They were then trained on the alternate interface (smartphone or embedded) for the same set of secondary tasks. Assessment of the alternate interface then took place during the second half of the drive as participants proceeded north on I-495. Most participants took approximately 35 to 40 minutes to drive each segment (north and south) (70 to 80 minutes combined).

The difficulty of the phone tasks was presented within each voice or manual period in the following order: easy, easy, hard, hard (see Section 2.3.1). This was intended to provide participants additional familiarity with the interface before assessing the harder task trials. Between individual trials, there was an interval of 30 seconds after the research associate recorded the completion of a task and the recorded instructions began for the next. A separation period of at least 3 minutes was provided following the end of one group of related tasks and the next period (e.g., between phone calling and address entry). Workload ratings for the second segment of the drive were completed along with a post-experimental questionnaire after completing the entire route. Total contact time for the study including intake and debrief was typically about 4 hours.

2.6 Dependent measures

2.6.1 Self-reported workload

Subjective workload ratings were obtained using a single global rating per secondary task type on a scale consisting of 21 equally spaced dots oriented horizontally along a 10 cm line with the numbers 0 through 10 equally spaced below the dots and end points labeled "Low" and "High" on the left and right, respectively. The rating scales for all the secondary tasks were presented on one sheet, allowing participants to rate the tasks relative to each other. Participants were instructed to "circle a point along each scale that best corresponds to how much workload you felt was involved in trying to do each task. Workload is best defined by the person doing the task and may involve *mental effort*, the amount of *attention required, physical effort, time pressure, distraction, or frustration* associated with trying to do the task while continuing to drive safely." Previous work (Beckers et al., 2014; Dopart et al., 2013) using this approach produced ratings across user interface tasks that were consistent with relative rankings obtained concurrently using the NASA-Task Load Index (Hart & Staveland, 1988; Hart, 2006), one of the most widely used subjective workload assessment scales. In this regard, Hendy, Hamilton, and Landry (1993) provide a useful consideration of the sensitivity of simple univariate workload scales relative to multifactor scales when the goal is to obtain an overall workload rating.

2.6.2 Task completion time

The time it takes to complete a task has often been used as one measure of demand/usability (e.g., Hornbæk & Law, 2007; Nielsen & Levy, 1994). Task duration has been considered in the automotive literature, particularly within the context of navigation entry tasks in evaluation of driver-interface usability and safety (e.g., Green, 1994, 1999a). Existing industry guidance in the form of SAE standard J2364 (SAE, 2004) recommends a maximum total task time of 15 seconds under static testing conditions for driver information systems that incorporate manual controls and visual displays; the guidelines explicitly state that this does not apply to voice-activated controls. The manner in which task duration is best interpreted in the context of voice-involved systems is largely an open question. In the current study, task completion time was calculated as the time between the end of a prompt to begin a task and completion of the task, which could involve successful completion (e.g., participant uttered the command to initiate a call to the specified contact), unsuccessful completion (e.g., participant uttered the command to initiate a call when it was not the specified contact), or failure at the point where the experimenter halted the trial (e.g., participant attempted to restart the full task for a third time, participant continued to pronounce an entry the same way multiple times and voice recognition could not interpret or misinterpreted).

2.6.3 Physiological metrics

Physiological metrics have long been used as objective measures of task demand in fields such as aviation (Kramer, 1991; Mulder, 1992; Roscoe, 1992; Veltman & Gaillard, 1998; Wilson, 2002) and have increasingly been employed in automotive-related research (Brookhuis & de Waard, 2001; Collet, Salvia, & Petit-Boulanger, 2014; Engström, Johansson, & Östlund, 2005; Lenneman & Backs, 2010; Mehler, Reimer, & Coughlin, 2012; Solovey, Zeck, Garcia-Perez, Reimer, & Mehler, 2014). Mehler, Reimer, Coughlin, and Dusek (2009) explored a range of peripheral physiological measures for differentiating objectively scaled levels of cognitive demand under driving simulation and found heart rate and skin conductance level (SCL) to be sensitive to task demand and practical to record. The same pattern of response was later observed for these two measures during actual highway driving (Reimer & Mehler, 2011) and their sensitivity was further characterised (Mehler et al., 2012). Heart rate and SCL were thus selected for inclusion in the current study.

The locations of R-wave peaks in the EKG signal were used to determine inter-beat intervals and calculate instantaneous heart rate using software developed at the MIT AgeLab. In line with existing standards (Task Force, 1996), automated detection results were visually reviewed and misidentified and irregular intervals manually corrected. Another MIT-developed data processing package removed high-frequency noise in the skin conductance signal, following Reimer and Mehler (2011), and identified motion artifacts were manually edited.

2.6.4 Visual metrics

The mean duration of individual (single) glances, the percentage of glances per participant greater than 2.0s, and the total time a participant glanced away from the forward road scene during a task were used as the primary glance metrics. These are the metrics specified by NHTSA (2013) in its recommended guidelines for evaluating the visual distraction associated with in-vehicle visual-manual electronic devices. The "away from the forward road scene" definition means that glances to other driving-relevant locations such as the instrument cluster and the rear and side mirrors are counted as looking away from the forward road scene. Prior to the release of NHTSA's guidelines, a more typical approach to task demand assessment had been to consider only glances to locations functionally relevant to the task under evaluation, such as a display or controls on the instrument cluster (e.g., Driver Focus-Telematics Working Group, 2006). Glance behavior was categorized as part of this study to support both characterizations. While the values obtained with each method differ and are worthy of consideration (e.g., Dopart et al., 2013; Reimer et al., 2013), using one or the other did not appreciably change the relative pattern of results in this dataset comparing systems and modalities. Given the potential relevance to ongoing guideline development, the metrics recommended by NHTSA are presented.

Eye orientation measures were quantified following ISO standards (ISO 15007-1, 2002; ISO 15007-2, 2001) with a glance to a region of interest defined to include the transition time to that object. In the manual coding of video images, the timing of glance is labeled from the first video frame illustrating movement to a "new" location of interest to the last video frame prior to movement to a "new" location. A recent analysis of driver glances to the center stack and other low-angle glances collected under the variable lighting conditions of real-world highway driving compared automated region of interest classification from commercial eye tracking equipment with video recordings (Reimer et al., 2013). The comparison found a high percentage of missing or inaccurate classifications in the automated output. Therefore, glance data for the current study were manually coded based on video of the driver following the taxonomy and procedures outlined in Reimer et al. (2013, Appendix G). Software, now available as open source (Reimer, Gruevski, & Couglin, 2014), allowed for rapid frame-by-fame review and coding. Each task period of interest was independently coded by two evaluators. Discrepancies between the evaluators (the identification of conflicting glance targets, missed glances, or glance timings that differed by more than 200ms) were mediated by a third researcher (see Smith, Chang, Glassco, Foley et al. (2005) for a discussion of the importance of multiple coders).

2.6.5 Vehicle control metrics

The vehicle control metrics were changes in vehicle speed, standard deviation of speed, and steering wheel reversal rates.

2.6.5.1 Vehicle speed. A reduction in mean vehicle speed (forward velocity) has frequently been observed during periods of increased task demand, and is often interpreted either as an attempt to increase safety margins or to reduce/manage the concurrent demands of the primary driving task and a secondary task (Angell et al., 2006). Speed reduction during secondary tasks tends to be more apparent when the task requires drivers to take their eyes off the forward roadway and/or actively involves taking a hand off the steering wheel, such as occurs when a driver holds a mobile phone to the ear (Brookhuis, de Vries, & de Ward, 1991; Engström, Johansson, & Östlund, 2005; Green, 1994; Patten, Kircher, Östlund, & Nilsson, 2004). Nominal increases in mean speed have sometimes been observed during pure auditory-vocal tasks such as conversing on a hands-free phone (Patten et al., 2004).

2.6.5.2 Standard deviation of vehicle speed. Variability in speed can be influenced by a range of factors, such as changes in the roadway environment and interactions with other drivers. To the extent that road conditions remain relatively constant, increased variability in speed can be interpreted as a reduction in direct attention to vehicle control and has been used at various times as a measure of control and/or changes in driver workload associated with secondary tasks (Green, 1994; Noy, 1990; Östlund et al., 2005).

2.6.5.3 Steering wheel reversal rates. During normal driving conditions, drivers typically make small steering wheel corrections to adjust vehicle heading for variations in roadway conditions (Liu, Schreiner, & Dingus, 1999). When visual attention is diverted from the roadway ahead, a driver's ability to make modest tracking responses is generally suspended until visual orientation to the roadway is regained. This results in periods of fixed steering wheel angle (Godthelp, Milgram, & Blaauw, 1984) and the need to make larger corrections upon return of the eyes to the roadway. Similarly, taking a hand off the steering wheel to operate a secondary control can result in more marked adjustments in steering. Östlund et al. (2004) found that visually demanding secondary tasks often invoke relatively large steering reversals of 2-6°, findings that were replicated in Engström, Johansson, and Östlund (2005). It is appropriate to note that it has been argued that steering wheel reversal rate is not a simple function of secondary task demand, but rather involves a complex interaction between primary task demand, secondary task demand(s), and the effort invested in the different tasks. McDonald and Hoffman (1980) suggested that steering frequency measures such as steering wheel reversal rate can reflect control effort and are not just a measure of tracking performance.

For purposes of this evaluation, *major steering wheel reversals* were considered as a control metric and classified as proposed in the final report of the European Union AIDE project (deliverable D2.2.5, section 7.12) (Östlund et al., 2005). This metric captures the number of steering wheel inputs exceeding an angular reversal gap of 3°. The rate of steering wheel reversals per minute was obtained by dividing the raw reversal rate by the task trial duration.

2.7 Data reduction & analysis

2.7.1 Subjective workload, behavioral and physiological measures

Baseline driving reference periods consisted of 2 minutes of just driving prior to a recorded audio message indicating that a new task period was about to start (see Figure 1). There were six such baseline periods per participant on the I-495 portion of the drive, and a seventh 2-minute reference was recorded on I-93 south on the return to MIT (14 minutes total). Values for relevant metrics were calculated, and the mean values across the baseline periods were used as an overall baseline/"just driving" reference. As already described, two trials of each type of phone calling and three trials of address entry using an embedded vehicle interface were presented to each participant. Values for each dependent measure were calculated per trial and mean values across trials were used for analytic purposes. All trials with usable data were included regardless of whether user or system errors occurred (see section 2.7.2 for more detail on error states and how they were handled). Trials with errors were included in the analysis as this was seen as more representative of the actual user experience than only considering error-free trials.

An evaluation of mean speed based on vehicle CAN data indicated a significant overall difference in speed between the two vehicles (Volvo M = 107.5 km/hr; Chevrolet M = 111.6 km/hr; $F(1,78) = 5.4$, $p < .023$); this difference was apparent even during baseline just driving periods (109.4 and 113.1 km/hr, respectively; $F(1,78) = 8.9$, $p = .004$). However, there was no significant difference in overall speed when considering data from the GPS units installed in each vehicle (Volvo M = 110.2 km/hr; Chevrolet M = 108.5 km/hr; $F(1, 78) = 0.53$, $p = .468$). This suggests that the CAN values may have systematically underestimated actual vehicle speed in the Volvo and overestimated speed in the Chevrolet. As a result, speed data from the GPS units were normalized as percentage changes relative to baseline driving periods for purposes of comparing the effects of interaction with the embedded systems in each vehicle. For consistency, GPS-based values were used for considering changes in standard deviation of speed as well.

Major steering wheel reversal rates were markedly higher during baseline driving in the Chevrolet (M = 20.39 per minute, SE = 0.9) versus the Volvo (M = 3.29, SE = 0.2) ($F(1,78) = 234.1$, $p < .001$). This could reflect basic tuning of the steering, other handling characteristics of the two vehicles, and/or differences in the quantification of steering wheel angle on the respective CAN buses. Consequently, comparisons of steering wheel metrics are reported considering percent changes relative to baseline driving.

Statistical analyses were performed in R (R Core Team, 2014) and an alpha level of 0.05 was used for assessing statistical significance. Owing to the non-normal distribution of the data and/or the use of ratio data (percentages) for several dependent measures, in many cases non-parametric statistics such as the Wilcoxon signed rank test and the Friedman test were used (similar to the t-test and repeated-measures ANOVA, respectively). For multifactorial analyses, repeated-measures ANOVA by ranks are presented. These tests have been shown to be more robust against Type I error in cases where data are non-normal (Conover & Iman, 1981; Friedman, 1937).

There were substantive differences between the contact calling and address entry tasks. For example, independent periods of contact calling with voice commands and manual entry were considered, while address entry was undertaken with voice commands only. Consequently, separate analyses were conducted for the two types of secondary tasks. The design for the contact calling was a mixed design with vehicle and the associated embedded system as a between-subject variable (MyLink or Sensus). There were two within-subject factors, modality (manual entry or voice entry), and task difficulty (easy or hard), resulting in a 2×2×2 mixed design. The full model was used in the analysis of the self-reported workload and task completion time data where effects for the easy and hard categorizations were of most interest for characterizing system implementation differences. Task difficulty was dropped from the model for analysis of physiology, eye glance, and driving performance metrics as typical use of the technologies would likely involve a mix of the easy and hard categories of contact calling. The analysis for the address entry task considered only embedded system (MyLink or Sensus) as a between-subject factor. Where applicable, tests comparing differences on selected variables between baseline driving and periods with the phone calling and navigation tasks are presented.

2.7.2 Error analysis & interaction characterization

A multi-step analysis of participants' interactions with the vehicle systems was carried out. The first analysis considered for each individual task trial whether it was error free or if a system or user-based error occurred. An example of a user error is a participant speaking an incorrect command during a voice-entry task that resulted in the task moving forward incorrectly or not moving forward at all. An example of a system error is the system misinterpreting a voice command that was in the correct form and understandable to the research associate present in the vehicle or a staff member listening to an audio recording of the interaction. Two members of the research staff independently evaluated each trial for errors (the research associate observing the participant during the drive and a second staff member who reviewed video and audio recordings of the interaction). A third staff member mediated any discrepancies. For the binary classification of whether a user or system error occurred during a trial, it was decided that if a user error and system error occurred in the same trial, the trial would be coded as a user error regardless of the number of each error type in the same trial. Thus, it is likely that the rate of system errors is underrepresented in this analysis.

The second error analysis was a more fine grained characterization of the extent to which participants experienced any difficulty in completing a task. Individual trials were classified as: 1) being completed without error or backtracking, 2) completed with backtracking, 3) completed with one instance of the research associate providing a prompt to assist the participant, 4) completed with more than one prompt from the research associate, or 5) failure to complete the task. The "backtracking" category covered situations where, for example, the system did not recognize or misinterpreted a street name, but the system dialog asked for confirmation and allowed for another opportunity for entry without exiting and requiring the participant to begin the entire task from the start. In other words, a backtracking classification indicates that the system successfully supported error recovery (arising from either user error or system recognition error). Backtracking could also occur because a participant recognized that they made an error (such as giving a wrong street name) and used an option provided by the system to correct the error. If the research associate judged that a participant was not going to progress through a task on their own, one or more limited prompts was provided to the participant. The intent here was largely to provide the participant with further assistance in learning how to use the system so that they might gain additional familiarity and become more successful on subsequent trials. If a participant had to restart a task more than twice or otherwise failed to progress at a point in the interaction despite support from the research associate, then the research associate guided the participant through terminating the trial and moved-on. Failure to progress could be due to either user or system errors. Trials that were terminated or that failed to progress due to either user or system errors were categorized as a failure.

3. Results

Beyond the 80 participants considered in the age and gender balanced analysis sample, there were a number of task relevant exclusions. These included eight individuals who were not taken on-road: two who experienced consistent voice recognition problems with a vehicle voice system (both with the Chevrolet MyLink); two who expressed discomfort with the idea of engaging in one or more of the tasks while driving after being exposed to them during training (both female, 64 years of age); and four who experienced significant difficulty trying to learn tasks in the parking lot (all male, 45-64 years). Of individuals taken on-road, exclusions included: one (63 year-old male) who was consistently unable to recall the actions necessary to complete tasks, requiring continuous prompting by the research associate; two (56 and 65 year-old males) for whom the research associate discontinued presentation of one or more task sets due to concerns on the research associate's part regarding the participant's ability to engage in tasks safely while driving. Other non-task relevant exclusions included three participants who were withdrawn during the drive due to broader safety concerns (one expressed drowsiness while driving, one frequently drifted out of lane, one with other unsafe driving habits) and four cases where weather and/or traffic conditions precluded continuing.

Findings for the analysis sample are presented first for the phone calling tasks followed by results for destination address entry tasks. Each section considers participants' subjective assessment of the workload associated with each task followed by objective data that include task duration, physiological measures, glance behavior, vehicle control metrics, and secondary task errors.

3.1 General sensitivity of physiological and driving metrics to secondary task periods

Changes in physiological arousal are characterized for analysis purposes as percentage changes relative to baseline driving to account for the different base values of individual participants. As expected, engaging with the secondary tasks while driving was associated with a higher state of arousal. Relative to baseline driving, there was on average an increase during the phone tasks across modalities and systems in heart rate of 2.2% (SE = 0.8) (W = 2516, $p < .001$) and an increase in skin conductance level of 11.3% (SE = 3.1) (W = 2397, $p < .001$). During the voice-command-based destination address entry across both systems, heart rate increased on average 1.5% (SE = 0.5) (W = 2112, p = .018) and skin conductance levels increased 7.3% (SE = 2.4) (W = 1956, p = .002).

Mean speed decreased significantly across the combined manual and voice-based phone calling tasks periods (*M* = -2.5%, *SE* = 1.1; *W* = 680, $p < .001$) although not during the voice-command-based address entry task periods (*M* = -0.4%, *SE* = 0.6; *W* = 1510, *p* = .559). Standard deviation of speed decreased across the manual and voice-based phone calling tasks (*M* = -37.6%, *SE* = 5.2; *W* = 76, $p < .001$) and the voice-based destination entry tasks (*M* = -19.9%, *SE* = 5.0; (*W* = 584, $p < .001$). The rate of major steering wheel reversals increased significantly across the combined manual and voice-based phone calling tasks (*M* = 31.9%, *SE* = 5.0; *W* = 2585, $p < .001$) but not during voice-based address entry (*M* = -0.49%, *SE* = 4.2; *W* = 1494, *p* = .547).

3.2 Phone Contact Calling

In considering the phone calling tasks, 'modality' refers to the overt method of interface interaction (manual or voice) and 'difficulty' refers to the easy or hard form of the task. Table 2 provides the means and standard errors of the measures used for analysis of the contact calling tasks presented by modality and embedded system type (Chevrolet MyLink or Volvo Sensus). An expanded set of tables providing details on measures not directly used in the analysis, such as alternate glance metrics, is provided in Table 5 and Table 6 in the Appendix.

Table 2: Means (and standard errors) by phone calling task and embedded vehicle system (Cheverolet MyLink or Volvo Sensus) for measures used for analysis.

	Vehicle	Phone Easy (Manual)	Phone Hard (Manual)	Phone Easy (Voice)	Phone Hard (Voice)
Self-Reported Workload	Chevrolet	4.28 (0.4)	5.20 (0.4)	1.81 (0.3)	1.90 (0.3)
	Volvo	5.49 (0.4)	6.12 (0.4)	2.05 (0.2)	2.55 (0.3)
Task Completion Time	Chevrolet	29.18 (2.0)	23.30 (0.9)	20.48 (1.3)	22.74 (1.8)
	Volvo	31.43 (2.0)	34.36 (2.6)	34.48 (1.3)	41.87 (1.6)
% Change in Heart Rate	Chevrolet	2.54 (0.9)	1.12 (0.8)	2.46 (0.8)	3.75 (0.8)
	Volvo	2.07 (1.0)	2.10 (0.7)	2.47 (0.7)	0.97 (0.6)
% Change in SCL	Chevrolet	15.06 (3.8)	13.15 (3.0)	13.66 (3.3)	12.22 (3.2)
	Volvo	13.09 (3.0)	12.75 (2.9)	7.62 (2.8)	3.63 (2.7)
Mean Off-Road Glance Duration	Chevrolet	0.89 (0.0)	0.94 (0.0)	0.60 (0.0)	0.61 (0.0)
	Volvo	0.94 (0.0)	0.95 (0.0)	0.79 (0.0)	0.79 (0.0)
% of Off-Road Glances > 2.0s	Chevrolet	1.73 (0.6)	3.12 (0.8)	0.09 (0.1)	0.49 (0.5)
	Volvo	2.98 (1.0)	3.80 (1.0)	0.94 (0.4)	0.57 (0.2)
Total Off-Road Glance Time	Chevrolet	15.16 (1.2)	11.97 (0.7)	3.42 (0.5)	3.23 (0.4)
	Volvo	15.95 (1.1)	16.82 (1.4)	9.78 (0.7)	10.65 (0.9)
Number of Off-Road Glances	Chevrolet	16.74 (1.1)	12.82 (0.6)	5.26 (0.7)	5.05 (0.6)
	Volvo	16.96 (1.0)	17.70 (1.3)	12.44 (1.0)	13.46 (1.0)
% Change Speed (GPS)	Chevrolet	-3.64 (1.6)	-4.63 (1.9)	-1.09 (1.0)	-0.13 (1.1)
	Volvo	-4.03 (0.9)	-2.14 (0.6)	-1.72 (0.7)	-2.56 (1.1)
% Change in SD of Speed (GPS)	Chevrolet	-33.48 (5.6)	-41.58 (4.3)	-54.76 (4.3)	-53.90 (3.6)
	Volvo	-27.36 (9.4)	-36.24 (3.8)	-30.57 (4.2)	-22.80 (6.3)
% Change in Major Wheel Reversals	Chevrolet	23.26 (8.4)	27.58 (7.2)	28.63 (9.2)	16.80 (10.5)

Volvo	51.41 (13.4)	38.11 (16.0)	31.31 (17.0)	37.84 (14.8)

3.2.1 Self-reported workload

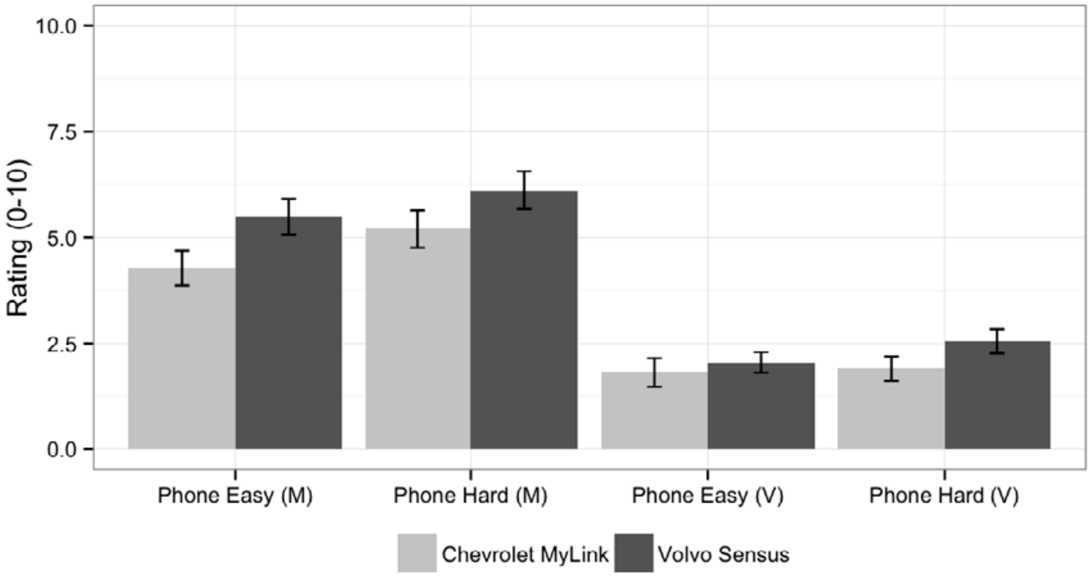


Figure 2: Mean self-reported workload ratings for all phone tasks by modality (manual or voice) and embedded system type (Chevrolet MyLink or Volvo Sensus) on a 0 (low) to 10 (high) scale. Error bars represent ±1 standard error.

A full breakdown of means and standard errors for self-reported workload by modality (manual or voice), system (Chevrolet MyLink or Volvo Sensus), and task difficulty (easy vs. hard) are presented in Figure 2 and Table 2. Self-reported workload for phone calling differed significantly by modality ($F(1,76) = 144.1$, $p < .001$) and difficulty level ($F(1,76) = 32.9$, $p < .001$). Mean ratings were higher for manual phone calling ($M = 5.3$, $SE = 0.40$) than for voice-based calling ($M = 2.1$, $SE = 0.28$). On average, the hard phone calling task had higher workload ratings than the easy phone calling task (easy $M = 3.4$, $SE = 0.33$; hard $M = 3.9$, $SE = 0.35$). On average, subjective workload ratings were lower with MyLink ($M = 3.3$, $SE = 0.35$) compared with the ratings with Sensus ($M = 4.0$, $SE = 0.33$); however, this difference only approached statistical significance ($F(1,76) = 3.31$, $p = .07$). There were no significant interactions between embedded system, modality, or task difficulty ($p > .16$).

3.2.2 Task completion time

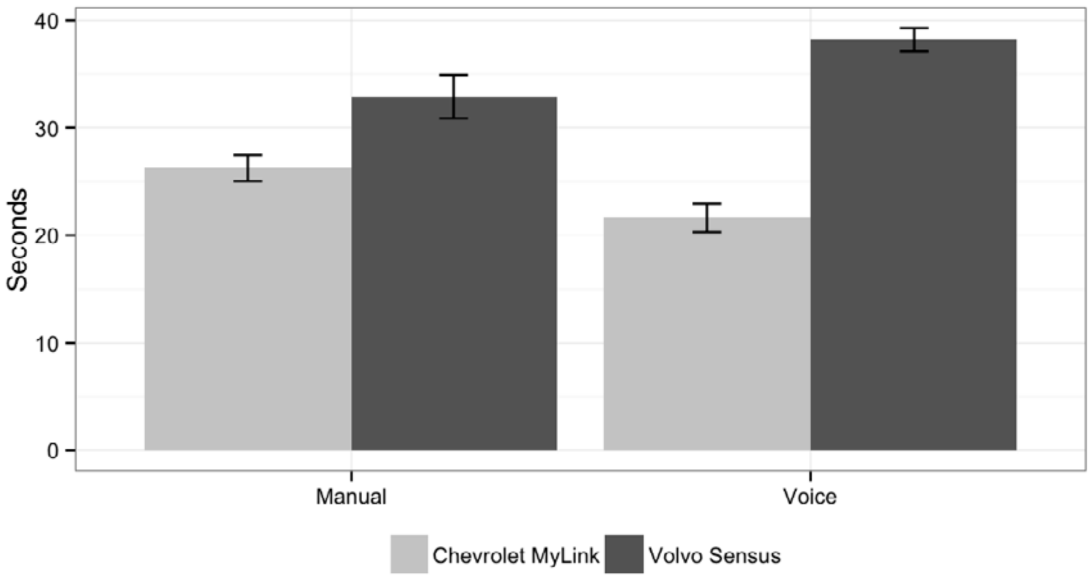


Figure 3: Mean completion time for phone calling by modality and type of embedded system. Error bars represent ±1 standard error.

In contrast with self-reported workload, there was no main effect of modality on task completion time ($F(1, 78) = 1.14$, $p = .288$), but there was a significant difference between the systems ($F(1,78) = 89.9$, $p < .001$) and a significant interaction between modality and system ($F(1,78) = 37.6$, $p < .001$) (see Figure 3). On average, participants using MyLink took longer to complete the phone calling task using the manual interface ($M = 26.2$ s, $SE = 1.5$) than the voice interface ($M = 21.6$ s, $SE = 1.6$). Conversely, participants using Sensus took longer using the voice interface ($M = 38.2$ s, $SE = 1.5$) than the manual interface ($M = 32.9$ s, $SE = 2.3$).

Significant interactions between modality and difficulty ($F(1,78) = 15.71$, $p < .001$) and between system and difficulty ($F(1,78) = 12.4$, $p < .001$) were present (see Table 2). Compared with the easy phone calling task, the hard phone calling task took longer to complete when participants used the voice interfaces (easy $M = 27.5$ s, $SE = 1.3$; hard $M = 32.3$ s, $SE = 1.7$) but took less time when using the manual interfaces (easy $M = 30.3$ s, $SE = 2.0$; hard $M = 28.8$ s, $SE = 1.8$). The task completion times for the easy phone calling and hard phone calling tasks were similar with MyLink ($M = 20.5$ s, $SE = 1.3$ and $M = 22.7$ s, $SE = 1.8$, respectively), but the hard phone calling task took 21 percent longer to complete than the easy phone calling task with Sensus (41.9 s and 34.5 s, respectively).

3.2.4 Physiological measures

There were no significant main effects of modality, system, or modality by system interaction across the tasks in terms of percentage change in either heart rate or skin conductance level during task periods relative to baseline (all $p > .05$). There was a three-way interaction between modality, system, and task difficulty for percentage change in heart rate ($F(1,78) = 15.2$, $p = .002$). The percentage change in heart rate among participants who used the MyLink voice interface to complete the hard contact calling task ($M = 3.75\%$, $SE = 0.8$) was greater than that for participants who used the voice interface in Sensus (M = 0.97%, $SE = 0.6$). The difference in percentage change in heart rate between MyLink and Sensus was not observed for the easy contact calling task using the voice interfaces or the easy or hard contact calling tasks with the manual interfaces.

3.2.5 Glance behavior

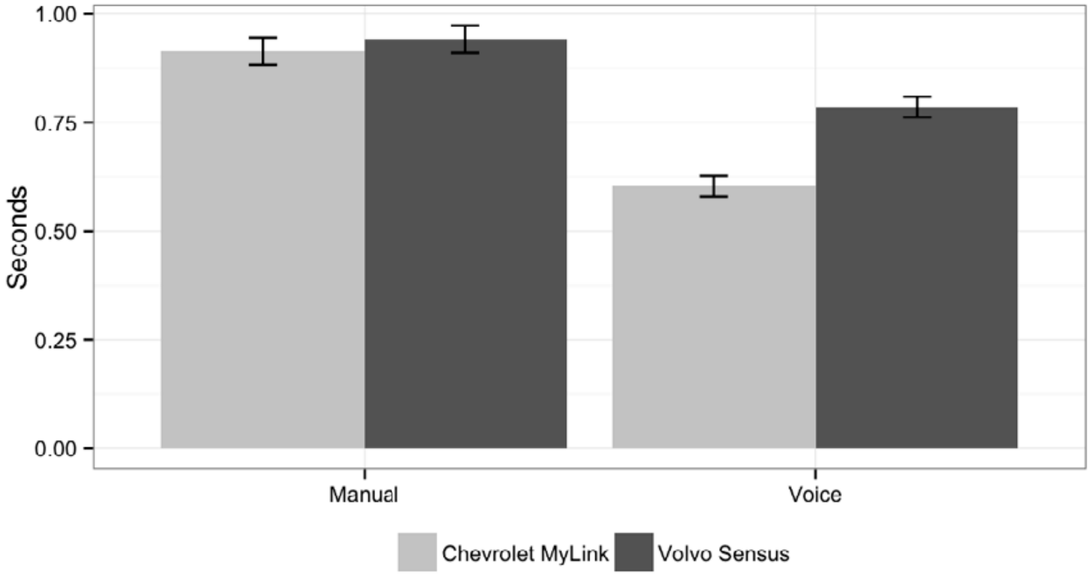


Figure 4: Mean single glance duration for all off-road glances during task periods by interface type and embedded system type. Error bars represent ±1 standard error.

There were significant main effects of modality ($F(1,78) = 204.8$, $p < .001$) and system $F(1,78) = 10.6$, $p = .002$), and a significant interaction between modality and system ($F(1,78) = 24.5$, $p < .001$) for mean single glance duration. Mean single glance duration for off-road glances was shorter when phone calling was performed using a voice interface ($M = 0.69$ s, $SE = 0.02$) compared with the manual interface ($M = 0.93$ s, $SE = 0.02$) in both vehicles; however, the reduction in single glance duration during voice interaction compared with manual interaction was greater with MyLink than Sensus (see Figure 4).

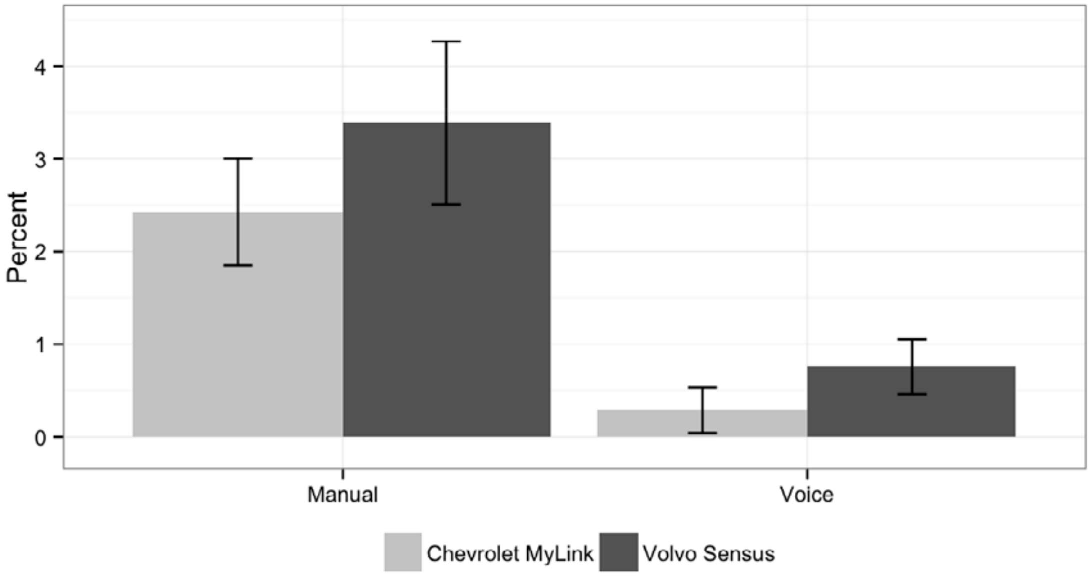


Figure 5: Percent of off-road glances greater than 2 seconds in duration. Error bars represent ±1 standard error.

On average, only a small percentage of participants' glances were longer than 2 seconds (*M* = 1.4%, *SE* = 0.3) (see Figure 5). Nonetheless, there was a significant main effect of modality ($F(1,78) = 39.0$, $p < .001$). On average, the percentage of glances that were longer than 2 seconds for each participant was smaller when using the voice interfaces (*M* = 0.5%, *SE* = 0.2) compared with using the manual interfaces (*M* = 2.9%, *SE* = 0.5). There was no significant main effect of system type (F(1,78) = 2.1, p = .149) or interaction between modality and system (F(1,78) = .100, p = .768).

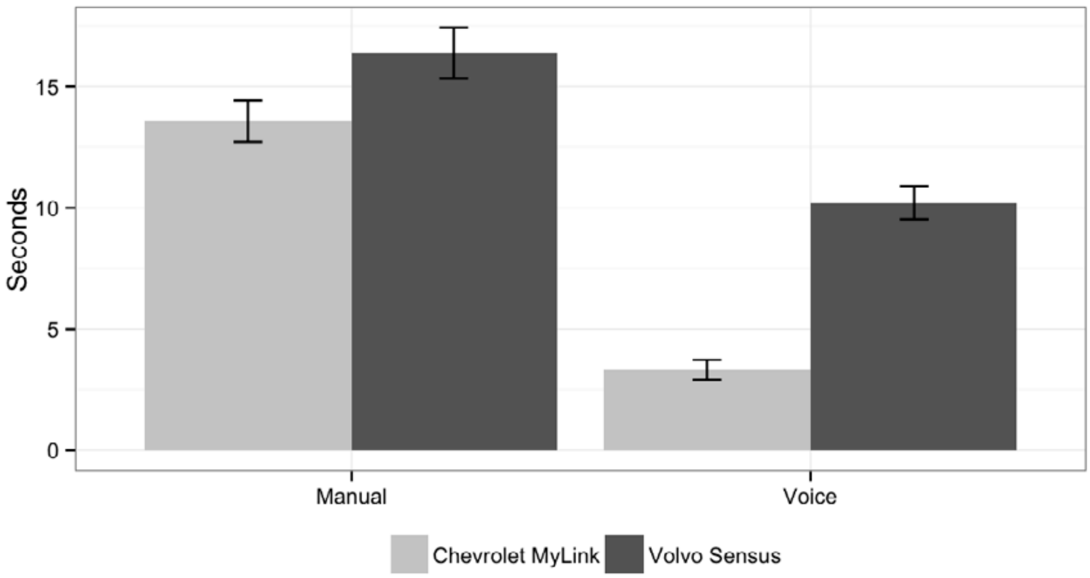


Figure 6: Total eyes-off-road time. Error bars represent ±1 standard error.

For total eyes-off-road time, there were significant main effects of modality ($F(1,78) = 266.8$, $p < .001$) and system ($F(1,78) = 35.3$, $p < .001$), and a significant interaction between modality and system ($F(1,78) = 30.6$, $p < .001$). The mean values for total eyes-off-road time were less when participants completed the phone calling task using the voice interfaces ($M = 6.8$ s, $SE = 0.6$) than when using the manual interfaces ($M = 15.0$ s, $SE = 0.7$); however, the reduction in total eyes-off-road time associated with using the voice interface relative to the manual interface was much greater for participants using MyLink than participants using Sensus (Figure 5). There was no significant main effect of task difficulty on total eyes-off-road time ($F(1,78) = 0.78$, $p = .379$).

3.2.6 Vehicle control metrics

As previously noted, on average, participants decreased their speed somewhat during the phone task periods. There was a greater percentage reduction in speed relative to baseline during manual phone calling ($M = 3.6\%$, $SE = 1.3$) compared with voice phone calling ($M = 1.4\%$, $SE = 1.0$) ($F(1,78) = 10.5$, $p = .002$). No significant main effect of system ($F(1,78) = 1.65$, $p = .202$) or interaction appeared ($F(1,78) = 0.22$, $p = .643$).

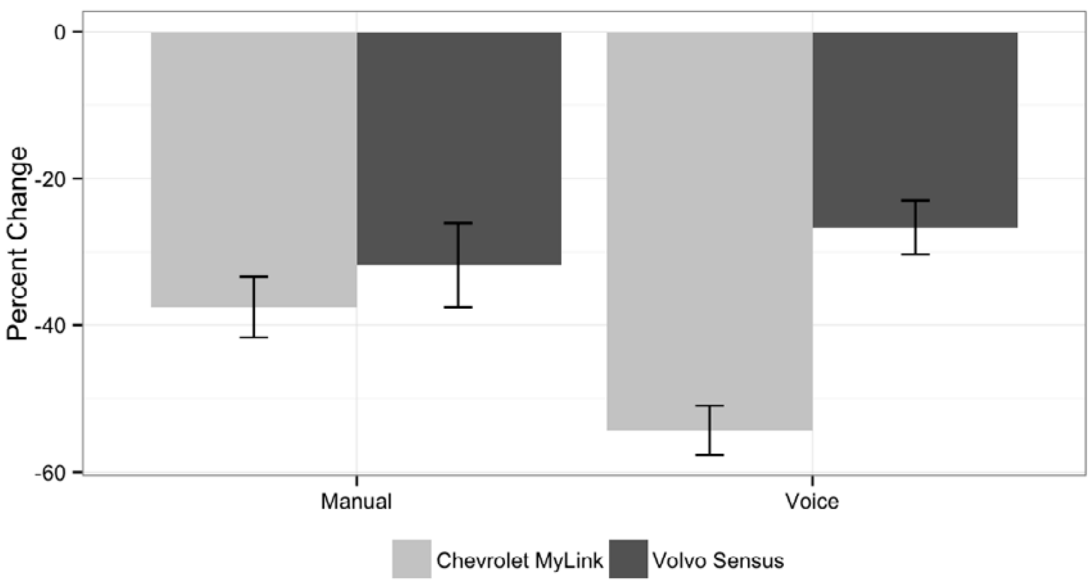


Figure 7: Mean percent change in standard deviation of speed (GPS) during phone task periods relative to baseline. Error bars represent ±1 standard error.

The was no overall main effect of modality on the percentage change in standard deviation of speed ($F(1,78) = 2.02$, $p = .159$). However, a significant main effect of system ($F(1,78) = 12.01$, $p = .001$) was present. On average, the reduction in standard deviation of speed during phone calling was greater for participants who used MyLink ($M = -45.9\%$, $SE = 4.5$) than for those who used Sensus ($M = -29.2\%$, $SE = 5.9$). In addition, a system by modality interaction was observed ($F(1,78) = 21.55$, $p < .001$) (see Figure 7). Detailing the interaction, there were no significant differences in the percentage reduction in standard deviation of speed between manual and voice calling with the Sensus (manual $M = -31.8\%$, $SE = 5$; voice $M = -26.7\%$, $SE = 5.3$) or between the two manual interfaces. In contrast, there were on average greater reductions in the standard deviation of speed with MyLink voice (*M* = -54.3%, *SE* = 4.0) than the manual (*M* = -37.5%, *SE* = 5.0) mode, and the reduction in the MyLink voice calling condition was greater than both Sensus conditions.

Overall, using the percentage change from baseline driving metric, the relative increase in major steering wheel reversals was nominally higher during manual calling (*M* = 35.1%, *SE* = 6.5) than during voice-command-based calling (*M* = 28.6%%, *SE* = 7.6); however, the difference was not statistically significant ($F(1,78) = 2.14$, $p = .148$). There was no significant main effect of system on major steering wheel reversal rates ($F(1,78) = 0.58$, $p = .45$) and no significant interaction between system and modality ($F(1,78) = 0.16$, $p = .69$).

3.3 Destination address entry into a navigation system

Descriptive statistics and analytic results considering the extent to which significant differences appeared between participant groups using the two voice command-based systems to enter destination addresses are presented in Table 3. An expanded listing including alternate eye glance metrics and absolute values for measures prior to conversion to percentage change scores appear in Table **7** in the Appendix.

Table 3 Means (and standard errors) and results of Wilcoxon signed rank tests for the destination address entry tasks. Change scores represent the percentage (%) change from baseline just driving.

	Chevrolet	**Volvo**	**W**	**P-value**	
Self-Reported Workload	3.59 (0.44)	2.54 (0.28)	924.5	0.154	
Task Completion Time	66.68 (2.85)	80.60 (1.71)	408	< 0.001	*
% Change in Heart Rate	1.66 (0.87)	1.25 (0.67)	801	0.996	
% Change in Skin Conductance Level	11.59 (3.77)	3.29 (2.44)	811	0.172	
Mean Off-Road Glance Duration	0.74 (0.02)	0.82 (0.02)	562	0.022	*
% of Off-Road Glances > 2.0sec	1.02 (0.29)	1.27 (0.36)	777.5	0.813	
Total Off-Road Glance Time	14.28 (1.22)	22.56 (1.43)	367	< 0.001	*
Number of Off-Road Glances	18.65 (1.52)	27.77 (1.75)	397	< 0.001	*
% Change in Speed (GPS)	0.60 (0.62)	-0.98 (0.55)	990	0.068	
% Change in SD of Speed	-29.53 (4.58)	-10.35 (5.46)	550	0.016	*
% Change in Major Wheel Reversals	7.34 (6.10)	-8.32 (6.82)	1003	0.051	

*p<.05

3.3.1 Subjective Workload

Mean self-reported workload for navigation address entry was nominally higher for the MyLink system (*M* = 3.59; *SE* = 0.44) than for Sensus (*M* = 2.54; *SE* = 0.28); however, this difference did not reach statistical significance ($W = 925$, $p = .15$).

3.3.2 Task Completion Time

There was a significant main effect of system on the time it took to complete the navigation address entry task ($W = 408$, $p < .001$). On average, participants using MyLink (*M* = 66.7 sec, *SE* = 2.85) completed the address entry task in less time than participants using Sensus (*M* = 80.6 sec, *SE* = 1.71).

3.3.3 Physiological Measures

While heart rate and SCL were higher during address entry than during baseline driving (see section 3.1), there was no significant effect of system for the percentage change in heart rate during address entry relative to baseline driving (MyLink *M* = 1.7%, *SE* = 0.9; Sensus *M* = 1.3%, *SE* = 1.7; *W* = 801, *p* = .996) or the percentage change in skin conductance level (MyLink *M* = 11.6%, *SE* = 3.8; Sensus *M* = 3.3%, *SE* = 2.4; *W* = 811, *p* = .172).

3.3.4 Glance Behavior

Mean single off-road glance duration during navigation address entry was significantly shorter for participants using MyLink (*M* = 0.74 sec, *SE* = 0.02) compared with participants using Sensus (*M* = 0.82 sec, *SE* = 0.02) (W = 562, *p* = .022). Similarly, the average total off-road glance time was significantly shorter for participants using MyLink (*M* = 14.3 s, *SE* = 1.2) than participants using Sensus (*M* = 22.6 s, *SE* = 1.4) (*W* = 367, *p* < .001). The overall number of long-duration glances was low, and there was no significant main effect of system on the percentage of glances made by a participant that were longer than 2 seconds (*W* = 777.5, *p* = .81).

3.3.5 Vehicle Control Metrics

The main effect of system on the percentage change in mean speed during navigation address entry relative to baseline approached statistical significance (*W* = 990, *p* = .068). Speed nominally increased among participants who used MyLink (*M* = 0.6%, *SE* = 0.62) but decreased for participants who used Sensus (*M* = -1.0%, *SE* = 0.55). Participants using MyLink showed a significantly greater reduction in their standard deviation of speed relative to baseline (*M* = -29.5%, *SE* = 4.6) than participants using Sensus (*M* = -10.4%, *SE* = 5.5) *W* = 550, *p* = .016). In terms of the percentage change in major steering wheel reversal rate relative to baseline driving, there was a nominal difference associated with system type during address entry (*W* = 1003, *p* = .051). The percentage change in major steering wheel reversal rate was 7.34% (*SE* = 6.1) for participants using MyLink and -8.3% (*SE* = 6.8) for participants using Sensus.

3.4 Errors & interaction characterization

Errors occurred in 7.3% of the phone calling trials (47 out of 640 trials). As can be observed in Table 4, errors attributable to a system were virtually nonexistent for manual contact calling (1 trial) and were present 2% of the time (7 trials) for voice-command entry. Considering both modalities together, user errors when attempting to call a contact were more prominent than system errors, occurring in 6.1% of the trials (*W* = 477.0, *p* < .001). If user and system errors are combined as a measure of usability and the two systems are considered together, no generalized advantage in frequency of trails with error appears by modality (manual: 23 trials; voice: 24 trials).

Table 4: Number of trials with errors and breakdown by type of error.

System	Task	Trials	Error Free	System Errors	User Errors	Total Errors
Chevrolet MyLink	Calling - Manual	160	153	0	7	7
Volvo Sensus	Calling - Manual	160	144	0	16	16
Chevrolet MyLink	Calling - Voice	160	147	5	8	13
Volvo Sensus	Calling - Voice	160	149	2	9	11
Chevroltet MyLink	Address Entry	120	59	38	23	61
Volvo Sensus	Address Entry	120	107	5	8	13

The overall rate of errors for voice-command-based entry of a destination address was markedly higher (30.8%) than the rate for voice-based phone calling (7.5%) (see Table 4). A significant difference by system was also apparent (W = 1128.5, p = .001). An error occurred in more than half of the address entry trials (51%) for MyLink compared with 10.1% for Sensus. Comparing error types, system-based errors represented a larger percentage of the errors for MyLink (38 out of 61 trials with errors; 62.3%) than Sensus (5 out of 13 trials with errors; 38.5%).

Given the much higher error rates for address entry, Figure 8 provides a characterization of the relative degree of difficulty participants experienced with each embedded navigation system in each of the three trials. It can be observed that only two outright failures to input the correct address occurred among participants using Sensus versus 24 failures experienced with MyLink. It can also be seen in Figure 8 that the address for trial 2 was more challenging to enter in both vehicles. Trial 3 consisted of the entry of each driver's own home address. It can be observed that during trial 3, 38 out of 40 participants were able to successfully enter their home address without external support using Sensus, while a more modest 27 out of 40 were successful using MyLink.

Figure 8: Characterization of participant experience by trial for destination address entry. The stacked scaling represents individual drivers sorted by their experience for an individual trial (i.e., 40 drivers per vehicle). “RA Assist” refers to prompting support provided by a research assistant as detailed in the methods.

4. Discussion

The findings for the embedded phone calling tasks add to previous research indicating that using voice interfaces to interact with an “infotainment” system can significantly reduce subjective workload and visual demand compared with using a manual interface. With both the Chevrolet MyLink and the Volvo Sensus embedded systems, participants reported significantly lower levels of subjective workload, had shorter mean single off-road glance durations, had fewer off-road glances longer than 2 seconds, and spent less time looking away from the forward roadway during voice-command phone calling compared with manual phone contact calling.

While participants assigned to both vehicles experienced a number of apparent advantages using voice commands relative to manual input for the embedded phone tasks, there are still potential trade-offs to be considered in evaluating net benefits and method of interaction more generally. For example, depending on the nature of the task and the implementation, voice-command interactions can take longer than using a manual interface. In an examination of radio tuning (Mehler et al., 2014; Reimer et al., 2013), manually pressing a radio preset button took less time than depressing a press-to-talk button and then verbally requesting a preset. In contrast, verbally requesting a specific station was faster and resulted in less diversion of the eyes from the roadway than making multiple button presses to change modes and frequency band and then manually rotating a tuning knob. Thus, a traditional manual interface seems to be more advantageous in the first case and the voice-command option more advantageous in the latter. The present study extends upon this level of detail by characterizing the extent to which system implementation differences can impact various variables. Consistent with the hypotheses that stemmed from Reagan and Kidd (2013), manual phone contact calling took more time than voice contact calling with the MyLink interface, whereas the opposite was true with the Sensus interface.

As is evident in the task completion time results, design philosophy and implementation differences in the voice-command-based systems can significantly impact objective metrics. Overall, the Sensus approach broke the task into discrete steps; this was most evident in the navigation system, which dealt with city, street name, and street number independently. In contrast, MyLink employed an initial “one-shot” approach in which the full address was presented in a single vocal string. With phone calling, the vocal string could be relatively simple (e.g., “Call Frank Scott at work.”), and this approach worked well for almost all participants. For address entry, however, results were quite different. When MyLink successfully parsed and decoded a one-shot full address string, the task was completed relatively quickly. However, a trade-off appears in a higher failure rate due to voice recognition errors by the system and user input errors. Only two outright failures in address entry occurred using Sensus, while 24 were recorded for MyLink in the analysis sample.

It is also worth considering the extent to which implementation characteristics outside of the fundamental voice recognition system design and capabilities might play a role in observed recognition errors. As detailed in a companion report (Reimer, Mehler, Reagan, et al, 2015), voice-recognition tasks in a dash-mounted smartphone also were evaluated in both vehicles. Although the same smartphone was being used, voice recognition errors were found to be higher in the Chevrolet Equinox than in the Volvo XC60. *Post hoc* sound level readings taken while the vehicles were traveling at 65 mph found that the Chevrolet had louder ambient noise levels than the Volvo in the 250 Hz (Chevrolet: 65 dBA; Volvo: 62 dBA) and 1000 Hz bands (Chevrolet: 62.6 dBA; Volvo: 60.1 dBA). Thus, one hypothesis to explain some of the variance in voice recognition errors might be the impact of ambient noise levels. This highlights the broader issue of system integration in automotive and other contexts, e.g. considering the optimization of a voice-system in the overall vehicle environment.

4.1 Training & mental models

In addition to voice recognition errors, some level of research staff prompting was required in a much higher percentage of cases during address entry while underway with the MyLink system, in spite of the fact that everyone was trained on the interface in the parking lot prior to going on-road. During the third trial, where participants were entering their home address, only one driver using Sensus required prompting to successfully complete the entry. For drivers working with MyLink, 7 needed prompting assistance to successfully complete the task.

It is possible that some of this performance differential may disappear if a user gains additional experience with a system. According to the subjective impressions of the research staff, a significant challenge for participants using the one-shot interface was learning to speak full addresses relatively rapidly and in a continuous stream, i.e., without pauses between a street name and the city name or long enunciation of individual digits of a street number. It appeared that "trying to help the system" by speaking slowly and with pauses between elements was, in fact, not a good strategy with this system. Designing systems that work with speech spoken in a relatively natural, continuous stream without pauses should ultimately benefit the consumer. However, this exemplifies the challenge of how to communicate functional design characteristics to novice users when they do not have a mental model for system operation or where their existing mental model does not match the implementation. It is likely that a frustrated user may limit, or discontinue altogether, use of a system that proves difficult to use initially. It is also plausible that better understanding of a system's model of operation would lead to more use and increase the potential advantages of using voice-based interfaces over manual interfaces. Further research could assess such a hypothesis in a longitudinal study.

4.2 Visual demands of voice interfaces

While the embedded voice command-based interfaces studied here were associated with lower visual demand than the embedded manual interfaces for phone calling, they were still highly multi-modal, including manual interactions and involving measurable time looking off the roadway. Visual engagement associated with a voice-command interface can vary markedly depending on the system design approach and the type of task. Total eyes-off-road time during voice-based phone calling was relatively brief, with a mean of around 3.3s for MyLink and a notably higher 10.3s with Sensus. During voice-based address entry, the mean total eyes-off-road time was 14.3s with MyLink and significantly longer at 22.6s with Sensus. Relatively long total eyes-off-road times were also observed during address entry in a 2010 Lincoln MKS system which employed a menu-based approach similar to the Sensus (Mehler et al., 2014; Reimer et al., 2013).

4.3 Cognitive demands

In addition to the visual demands documented here, the question of the extent to which cognitive demands are an issue in voice-command systems remains a valid and challenging question. Reagan and Kidd (2013) specifically note the concern that although voice interfaces reduce visual demand, secondary activities, regardless of input modality, may produce levels of cognitive demand that may reduce road users' safety compared with just driving. Studies have shown that increased cognitive demands result in more constrained visual scanning patterns (Recarte & Nunes, 2000; Reimer et al., 2012), suppression of brain activity in visual processing areas (Just et al., 2008), and degradation of vehicle control on the test track (Owens et al., 2011). Likewise, Lo and Green (2013) observed that voice interfaces have been shown to offer various advantages, but still require cognitive demand, which can interfere with the primary driving task. Strayer et al. (2013) provide a particularly extensive review of reasons why cognitive demands arising from auditory-vocal interactions with technologies could be problematic when driving.

Viewed broadly, the voice tasks studied here did not appear to produce high cognitive workloads compared with other secondary tasks studied previously (Mehler et al., 2014; Reimer & Mehler, 2013). Self-reported workload was lower for both voice-based phone calling and destination address entry than what was reported for manual phone calling. Considering physiological arousal as an indicator of workload, increases were present during all voice and manual tasks, but did not differentiate between modalities. Compared with data collected in Mehler et al. (2014), elevations in heart rate were in the same general range as that induced by the 0-back level of the n-back surrogate working memory task (generally considered a very low cognitive demand task) and skin conductance values were nominally below the 1-back level (generally considered a moderately demanding cognitive task). Thus, while demands with voice interaction were present in the current study, the findings may not warrant the degree of concern raised in recent evaluations of embedded voice systems (e.g., Strayer, Turrill, Coleman, Ortiz, & Cooper, 2014), particularly when considering the several measures that indicate lower demand for the two embedded voice systems tested here relative to their manual counterparts. The present work provides additional evidence in two different vehicle implementations that voice-based interfaces are multi-modal in nature, drawing upon auditory, vocal, visual, manipulative, and cognitive resources. At a minimum, the consideration of visual demand, a well-established key correlate to safety, must be taken into account in developing a comprehensive assessment of voice interfaces. It is clear that providing a voice-interface does not inherently mean that drivers will or can keep their eyes continuously on the road.

Nevertheless, the data collected in the current study do not exhaustively explore the extent to which drivers might become so absorbed in a secondary task that look-but-do-not-see events become an issue or that frustration over problematic voice recognition might divert attention. Well-developed work to better understand the extent of these issues is needed. In this context, comprehensive assessment of cognitive absorption in voice-involved interactions should be considered relative to purely visual-manual alternatives in addition to "just driving." For example, two simulation studies of smartphone interactions found that drivers took longer to notice a light stimulus and missed more of the stimuli overall when using voice-based entry of a destination address compared with baseline driving (Beckers et al., 2014; Munger et al., 2014). At the same time, response rates and miss values were significantly lower for the voice-based interactions than for interactions with the visual-manual interface. Thus, while it is important to recognize that voice-interfaces are not free of demands on attention, it is also important to better understand the relative risks of different types of interactions while driving and to communicate this understanding to the public.

4.4 Limitations

The data presented characterize the behavior of drivers who were trained on the use of the information systems tested. Compared with actual owners of a vehicle who use such systems regularly, the population of study had limited experience. Furthermore, their interaction with the systems was evaluated at designated points during a structured drive. It is unknown how such an experimental evaluation mirrors the manner in which drivers generally use such systems and the self-regulatory patterns that accompany secondary task engagement. It might reasonably be expected that driver performance and comfort could improve with additional experience and greater self-selection of the points at which they engage with the systems. The extent to which this would impact the relative demand profiles across the interface models and the systems observed here is unknown. On the other hand, compared with other novice users, participants were given an in-depth introduction to the systems, which included guidance on short-cut methods to accomplish the tasks, and participants who were taken on-road practiced with the systems in a parking lot until they indicated they understood how to use them. The extent to which other novice users would attempt to actually use the technologies on-road without similar training and context is unknown.

Throughout the research protocol, multiple instructions presented in written form, recorded audio, and verbal reinforcement by a research associate emphasized that participants need not engage in a secondary task if they felt unsafe or if they would not typically engage in the tasks during their personal driving. As previously detailed, two older participants expressed reservations during training about engaging with the tasks while driving and did not go on-road; four additional older participants had sufficient difficulty learning the tasks that the research associate declined to proceed to on-road assessment. No participants who went on-road declined to engage with a task. However, one older participant was unable to recall the training sufficiently or deduce operation of most of the tasks while underway. In the case of two other participants, task presentation was discontinued due to a research associate's concern over the participants' ability to engage with the tasks safely while underway. These cases were relatively equally distributed across the two vehicles and not included in the analysis set, but should be kept in mind in terms of broader usability considerations of the self-reported workload data and other variables presented.

While the data presented here shows that interaction with voice interfaces can involve substantial visual engagement, a direct connection to driving safety risk is difficult to establish. The type of data presented here is informative concerning the attentional demand characteristics of the interface tasks, rather than necessarily being predictive of risk to drivers who are operating their own vehicles. Additional naturalistic and/or epidemiological research will be required to evaluate the extent to which interactions with these embedded vehicle systems present any significant elevation in risk.

In the current study, the measures of cognitive demand were not exhaustive, and different measures might provide an alternate perspective. It is also possible that other voice-command implementations or interactions with the systems under study (e.g. without awareness of shortcuts) might be associated with greater or lesser overt levels of cognitive or visual demand.

The presentation sequence used for the easy and hard phone tasks could be seen as a methodological limitation. The hard tasks were intentionally presented last to provide participants with maximum exposure to the contact calling interfaces prior to assessing the hard tasks so as to reduce the effect of novelty on the most challenging task. While not considered in detail in the results presented here, several measures suggested that some learning took place over the initial trials of basic phone calling such that, in some instances, a presumed hard task appeared less demanding than the easy task. For example, total task time for manual phone calling in the hard phone task was lower than that observed for the earlier easy trials in the MyLink system.

5. Conclusions

The comparison of manual and voice phone calling with the 2013 Chevrolet Equinox MyLink system and the 2013 Volvo XC60 Sensus system indicates that auditory-vocal interfaces can provide drivers with a means to decrease the time that their eyes are drawn away from the forward roadway when engaging in this type of secondary task. As was found in previous on-road research with actual production systems (Chiang et al., 2005; Mehler et al., 2014; Reimer et al., 2013; Reimer, Mehler, Dobres et al., 2014), the embedded voice interfaces studied here significantly reduced mean single glance time, the percentage of long duration glances (> 2s), and total off-road glance time relative to embedded manual interfaces.

An important extension in the current study compared with prior research is the comparison of the impact of differing system design approaches. As anticipated based on the task analysis of Reagan and Kidd (2013), the streamlined 'one-shot' approach of MyLink showed a distinct advantage with regard to total task time and several visual demand metrics compared with the layered menu-based approach of Sensus. However, limitations in voice recognition and parsing technology were more apparent with MyLink, where one-shot entry of a full address much more frequently resulted in voice recognition errors by the system, as well as more user input errors and difficulty using the system without assistance. These errors may result in increased workload from frustration associated with repeated engagement or cessation of an engagement in favor of an alternative approach such as visual-manual interaction, use of a smartphone, etc. Similarly, the recent report by Strayer et al. (2014) suggests that significant differences in voice system demand can be observed across vehicles, while Reimer et al. (2014) and Munger et al. (2014) showed in an embedded vehicle system and smartphone that differences in demand can occur based upon system settings.

Taken as a whole, these findings suggest both support for and caution in the development of auditory-vocal interfaces for use by drivers. While a properly designed and used interface can significantly reduce eyes-off-road time, neither of the interfaces studied here eliminated visual demand. Further, overall task duration and visual engagement were quite extensive in the case of destination address entry when compared with the traditional reference of manual radio tuning (NHTSA, 2013). The complex relationship between the observed levels of visual engagement and the time involved with voice-based interactions requires further study. It is unclear to what extent risk-based guidance developed for visual demand during manual interactions are directly applicable to voice-based interfaces. 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. Clearly, evaluations that ignore the complex intertwining of resource demands placed upon the driver paint an incomplete picture of the benefits and limitations associated with various interface design approaches and implementations. Future work needs to further investigate how different interface designs manage the transitions between the auditory-vocal and visual-manual subcomponents of a voice-based activity.

6. Acknowledgments

Primary support for this work was provided by the Insurance Institute for Highway Safety (IIHS), with additional support provided through a grant by the US DOT's Region I New England University Transportation Center at MIT. Acknowledgment is gratefully extended to Peter Hamscher, Alex Hruska, Martin Lavallière, Alea Mehler, Hale McAnulty, Mauricio Muñoz, Lauren Parikhal, Anthony Pettinato, Adrian

Rumpold, and Andrew Sipperley for their contributions in protocol development, data collection, reduction, and manual scoring. Appreciation is also extended to Adrian Lund and David Zuby for review and comment on the manuscript.

References

Angell, L., Auflick, J., Austria, P. A., Kochhar, D., Tijerina, L., Biever, W., Diptiman, T., Hogsett, J., Kiger, S. (2006). Driver Workload Metrics Task 2 Final Report & Appendices. Final Research Report: DOT HS 810 635. NHTSA/USDOT, Washington, D.C.

Barón, A., & Green, P. (2006). Safety and usability of speech interfaces for in-vehicle tasks while driving: a brief literature review. Ann Arbor, MI: The University of Michigan Transportation Research Institute (UMTRI).

Beckers, N., Schreiner, S., Bertrand, P., Reimer, B., Mehler, B., Munger, D., & Dobres, J. (2014). Comparing the demands of destination entry using Google Glass and the Samsung Galaxy S4. *Proceedings of the 58th Annual Meeting of the Human Factors and Ergonomics Society*, Chicago, IL, Oct. 27 - 31, 2014.

Blanco, M., Biever, W.J., Gallagher, J.P., & Dingus, T.A. (2006). The impact of secondary task cognitive processing demand on driving performance. *Accident Analysis and Prevention, 38*, 895-906.

Brookhuis, K.A., de Vries, G., & de Ward, D. (1991). The effects of mobile telephoning on driving performance. *Accident Analysis and Prevention, 23(4),* 309–316.

Brookhuis, K. A., & de Waard, D. (2001). Assessment of drivers' workload: Performance and subjective and physiological indexes. In P. A. Hancock & P. A. Desmond (Eds.), *Stress, workload, and fatigue* (pp. 321–333). Mahwah, NJ: Lawrence Erlbaum.

Chiang, D.P., Brooks, A.M., & Weir, D.H. (2005). Comparison of visual-manual and voice interaction with contemporary navigation system HMIs (Paper no. 2005-01-0433). *SAE 2005 World Congress & Exhibition*. SAE International: Warrendale, PA.

Collet, C., Salvia, E., & Petit-Boulanger, C. (2014). Measuring workload with electrodermal activity during common braking actions. *Ergonomics*, *57*(6), 886-896.

Conover, W. J., & Iman, R. L. (1981). Rank transformations as a bridge between parametric and nonparametric statistics. *American Statistician*, *35*(3), 124–129.

Dopart, C., Häggman, A., Thornberry, C., Mehler, B., Dobres, J., & Reimer, B. (2013). A driving simulation study examining destination entry with iPhone iOS 5 Google Maps and a Garmin portable GPS system. *Proceedings of the 57th Annual Meeting of the Human Factors and Ergonomics Society*, San Diego, CA, Sept. 30 - Oct. 4, 2013.

Driver Focus-Telematics Working Group (2006). Statement of principles, criteria, and verification procedures on driver-interactions with advanced in-vehicle information and communication systems. Alliance of Automobile Manufacturers: Washington, DC.

Engström, J., Johansson, E., & Östlund, J. (2005). Effects of visual and cognitive load in real and simulated motorway driving. *Transportation Research Part F*, 8, 97-120.

Friedman, M. (1937). The use of ranks to avoid the assumption of normality implicit in the analysis of variance. *Journal of the American Statistical Association*, *32*(200), 675–701.

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. (Report No. DOT HS 811 757). Washington, DC: National Highway Traffic Safety Administration.

Garay-Vega, L., Pradhan, A.K., Weinberg, G., Schmidt-Nielsen, B., Harsham, B., Shen, Y., Divekar, G., Romoser, M., Knodler, M., & Fisher, D.L. (2010). Evaluation of different speech and touch interfaces to in-vehicle music retrieval systems. *Accident Analysis and Prevention, 42*, 913-920.

Godthelp, H., Milgram, P., & Blaauw, J. (1984). The development of a time-related measure to describe driving strategy. *Human Factors, 26(3),* 257–268.

Graham, R. & Carter, C. (2000). Comparison of speech input and manual control of in-car devices while on the move. *Personal Technologies, 4*, 155-164.

Green, P. (1994). Measures and methods used to assess the safety and usability of driver information systems. Report No. UMTRI-93-12./FHWA-RD-94-088.

Green, P. (1999a). The 15-second rule for driver information systems. In *Proceedings of the ITS America Ninth Annual Meeting*.

Green, P. (1999b). Estimating compliance with the 15-second rule for driver-interface usability and safety. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 43, No. 18, pp. 987-991). SAGE Publications.

Haigney, D.E., Taylor, R.G., & Westerman, S.J. (2000). Concurrent mobile (cellular) phone use and driving performance: task demand characteristics and compensatory processes. *Transportation Research Part F, 3*, 113-121.

Harbluk, J.L., Burns, P.C., Lochner, M., & Trbovich, P.L. (2007). Using the lane-change test (LCT) to assess distraction: Tests of visual-manual and speech-based operation of navigation system interfaces. *PROCEEDINGS of the 4th International Driving Symposium on Human Factors in Driver Assessment, Training, and Vehicle Design*, 16-22, Iowa City, IA: University of Iowa Public Policy Center.

Hart, S. G. (2006, October). NASA-task load index (NASA-TLX); 20 years later. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 50, No. 9, pp. 904-908). Sage Publications.

Hart, S. G. & Staveland, L. E. (1988) Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In P. A. Hancock and

N. Meshkati (Eds.) *Human Mental Workload*. Amsterdam: North Holland Press.

Hendy, K.C, Hamilton, K.M. & Landry, L.N. (1993) Measuring subjective workload: when is one scale better than many? *Human Factors, 35 (4),* 579-601.

Hornbæk, K., & Law, E.L.C. (2007). Meta-analysis of correlations among usability measures. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems* (pp. 617-626). ACM.

ISO 15007-1. (2002). Road vehicles - Measurement of driver visual behaviour with respect to transport information and control systems - Part 1: Definitions and parameters. Geneva, Switzerland: International Standards Organization.

ISO 15007-2. (2001). Road vehicles - Measurement of driver visual behaviour with respect to transport information and control systems - Part 2: Equipment and procedures. Geneva, Switzerland: International Standards Organization.

Just, M.A., Keller, T.A., & Cynkar, J. (2008). A decrease in brain activation with driving when listening to someone speak. *Brain Research, 1205*, 70-80.

Klauer, S.G., Dingus, T.A., Neale, V.L., Sudweeks, J.D., & Ramsey, D.J. (2006). *The impact of driver inattention on near-crash/crash risk: An analysis using the 100-car naturalistic driving study data* (DOT HS-810-594). Washington, DC: National Highway Traffic Safety Administration.

Kramer, A. E. (1991). Physiological metrics of mental workload: A review of recent progress. In D. L. Damos (Ed.), *Multiple-task performance* (pp. 279–328). London, UK: Taylor & Francis.

Kubose, T.T., Bock, K., Dell, G.S., Garnsey, S.M., Kramer, A.F., & Mayhugh, J. (2006). The effects of speech production and speech comprehension on simulated driving performance. *Applied Cognitive Psychology, 20*, 43-63.

Lee, J. D., Caven, B., Haake, S., Brown, T. L. (2001). Speech-based interaction with in-vehicle computers: The effect of speech-based e-mail on drivers' attention to the roadway. *Human Factors, 43(4),* 631–640.

Liu, Y. C., Schreiner, C. S., & Dingus, T. A. (1999). *Development of human factors guidelines for advanced traveler information systems (ATIS) and commercial vehicle operations (CVO): Human factors evaluation of the effectiveness of multi-modality displays in advanced traveler information systems* (No. FHWA-RD-96-150,).

Lenneman, J.K., & Backs, R.W. (2010, November). Enhancing assessment of in-vehicle technology attention demands with cardiac measures. In *Proceedings of the 2nd International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (pp. 20-21). ACM.

Lo, V.E-W., & Green, P.A. (2013). Development and evaluation of automotive speech interfaces: useful information from the human factors and the related literature. *International Journal of Vehicular Technology, 2013,* ID 924170. http://dx.doi.org/10.1155/2013/924170

Maciej, J. & Vollrath, M. (2009). Comparison of manual vs. speech-based interaction with in-vehicle information systems. *Accident Analysis and Prevention, 41*, 924-930.

McDonald, W. A., & Hoffman, E. R. (1980). Review of relationships between steering wheel reversal rate and driving task demand. *Human Factors 22*(6), 733-739.

Mehler, B., Reimer, B., & Coughlin, J.F. (2012). Sensitivity of physiological measures for detecting systematic variations in cognitive demand from a working memory task: An on-road study across three age groups. *Human Factors, 54*(3), 396-412.

Mehler, B., Reimer, B., Coughlin, J. F., & Dusek, J. A. (2009). The impact of incremental increases in cognitive workload on physiological arousal and performance in young adult drivers. *Transportation Research Record: Journal of the Transportation Research Board, 2138*, 6–12.

Mehler, B., Reimer, B., Dobres, J., McAnulty, H., Mehler, A., Munger, D., & Coughlin, J.F. (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.* MIT AgeLab Technical Report No. 2014-2. Massachusetts Institute of Technology, Cambridge, MA.

Mulder, L. J. (1992). Measurement and analysis methods of heart rate and respiration for use in applied environments. *Biological Psychology, 34*(2–3), 205–236.

Munger, D., Mehler, B., Reimer, B., Dobres, J., Pettinato, A., Pugh, B., & Coughlin, J.F. (2014). A Simulation Study Examining Smartphone Destination Entry while Driving. *Proceedings of the 6th International Conference on Automotive User Interfaces and Interactive Vehicle Applications (AutoUI 2014),* Seattle, WA, September 17-19, 2014.

National Highway Traffic Safety Administration (2013). *Visual-Manual NHTSA Driver Distraction Guidelines for In-Vehicle Electronic Devices (Docket No. NHTSA-2010-0053).* U.S. Department of Transportation National Highway Traffic Safety Administration (NHTSA), Washington, DC, 2013.

Nielsen, J., & Levy, J. (1994). Measuring usability: preference vs. performance. *Communications of the ACM*, *37*(4), 66-75.

Noy, Y.I. (1990). *Attention and Performance While Driving with Auxiliary In-Vehicle Displays* (Transport Canada Publication TP 10727 (E)), Ottawa, Ontario, Canada: Transport Canada, Traffic Safety Standards and Research, Ergonomics Division.

Östlund, J., Nilsson, L., Carsten, O., Merat, N., Jamson, S., Janssen, W., Mouta, S., Carvalhais, J, Santos, J., Anttila, V., Sandberg, H., Luoma, de Waard, D., Brookhuis, K., Johansson, E., Engström, J., Victor, T., Harbluk, J., Janssen, W., & Brouwer, R., (2004). Deliverable 2—HMI and safety-related driver performance. Human Machine Interface and the Safety of Traffic in Europe (HASTE) Project, Report No. GRD1/2000/25361 S12.319626

Östlund, J., Peters, B., Thorslund, B., Engström, J., Markkula, G., Keinath, A., Horst, D., Juch, S., Mattes, S., & Foehl, U. (2005). Adaptive Integrated Driver-Vehicle Interface (AIDE): Driving performance assessment - methods and metrics. (Report No. IST-1-507674-IP). Information Society Technologies (IST) Programme, Gothenburg, Sweden.

Owens, J. M., McLaughlin, S.B., & Sudweeks, J. (2010). On-road comparison of driving performance measures when using handheld and voice-control interfaces for mobile phones and portable music players (Paper No. 2010-01-1036). *SAE 2010 World Congress & Exhibition*. SAE International: Warrendale, PA.

Owens, J. M., McLaughlin, S. B., & Sudweeks, J. (2011). Driver performance while text messaging using handheld and in-vehicle systems. *Accident Analysis & Prevention, 43*, 939-947.

Patten, C. J. D., Kircher, A., Östlund, J., & Nilsson, L. (2004). Using mobile telephones: Cognitive workload and attention resource allocation. *Accident Analysis and Prevention, 36,* 341–350.

R Core Team (2014). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Ranney, T.A., Harbluk, J.L., & Noy, Y.I. (2005). Effects of voice technology on test track driving performance: Implications for driver distraction. *Human Factors, 47*(2), 439-454.

Reagan, I.J. & Kidd, D.G. (2013). Using heirarchical task analysis to compare four vehicle manufacturers' infotainment systems.*Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 1495-2599. Santa Monica, CA: Human Factors and Ergonomics Society.

Recarte, M.A. & Nunes, L.M., (2000). Effects of verbal and spatial-imagery tasks on eye fixations while driving. *Journal of Experimental Psychology Applied, 6*(1), 31–43.

Reimer, B., Gruevski, P., & Coughlin, J.F. (2014). MIT AgeLab Video Annotator, Cambridge, MA. https://bitbucket.org/agelab/annotator

Reimer, B., & Mehler, B. (2011). The impact of cognitive workload on physiological arousal in young adult drivers: A field study and simulation validation. *Ergonomics, 54*(10), 932–942.

Reimer, B & Mehler, B. (2013). *The effects of a production level "voice-command" interface on driver behavior: Summary findings on reported workload, physiology, visual attention, and driving performance* (White Paper 2013-18A). Cambridge, MA: Massachusetts Institute of Technology AgeLab. Accessed at http://web.mit.edu/reimer/www/pdfs/MIT_AgeLab_White_Paper_2013-18A_(Voice_Interfaces).pdf.

Reimer, B., Mehler, B., Dobres, J. & Coughlin, J.F. (2013). *The effects of a production level "voice-command" interface on driver behavior: reported workload, physiology, visual attention, and driving performance*. MIT AgeLab Technical Report No. 2013-17A. Massachusetts Institute of Technology, Cambridge, MA.

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. *Proceedings of the 6th International Conference on Automotive User Interfaces and Interactive Vehicular Applications (AutoUI 2014),* Seattle, WA, September 17-19, 2014.

Reimer, B., Mehler, B., Reagan, I, Kidd, D., & Dobres, J. (2015). Multi-modal demands of a smartphone used to place calls and enter addresses during highway driving relative to two embedded systems. Manuscript under review (*Ergonomics*).

Reimer, B., Mehler, B., Wang, Y., & Coughlin, J.F. (2012). A field study on the impact of variations in short term memory demands on drivers' visual attention and driving performance across three age groups. *Human Factors, 54*(3), 454-468.

Roscoe, A. H. (1992). Assessing pilot workload: Why measure heart rate, HRV and respiration. *Biological Psychology, 34*(2–3), 259–287.

SAE (2004). Navigation and route guidance function accessibility while driving (SAE J2364). SAE International, Warrendale, PA.

Schreiner, C. S., Blanco, M. & Hankey, J. (2004). Investigating the effect of performing voice recognition tasks on the detection of forward and peripheral events. *Proceedings of the Human Factors and Ergonomics Society 48th Annual Meeting*, Santa Monica, CA, September, 2004, pp. 2354-2358.

Shutko, J., Mayer, K., Laansoo, E., & Tijerina, L. (2009). Driver workload effects of cell phone, music player, and text messaging tasks with the Ford SYNC voice interface versus handheld visual-manual interfaces (Paper No. 2009-01-0786). *SAE 2009 World Congress & Exhibition*. SAE International: Warrendale, PA.

Smith, D.L., Chang, J., Glassco, R., Foley, J., & Cohen, D. (2005). Methodology for capturing driver eye glance behavior during in-vehicle secondary tasks. *Transportation Research Record: Journal of the Transportation Research Board, 1937(1),* 61-65.

Solovey, E., Zeck, M., Garcia-Perez, E.A., Reimer, B., & Mehler, B. (2014). Classifying driver workload using physiological and driving performance data: two field studies. *Proceedings of the 32nd Annual Conference on Human Factors in Computing Systems (CHI 2014),* Toronto, Canada, April 26 – May 1, 2014, pp. 4057-4066.

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.

Strayer, D.L., Drews, F.A., & Johnston, W.A. (2003). Cell phone-induced failulres of visual attention during simulated driving. *Journal of Experimental Psychology: Applied, 9*(1), 23-32.

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*. Washington, DC: AAA Foundation for Traffic Safety.

Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology (1996). Heart rate variability: standards of measurement, physiological interpretation, and clinical use. *European Heart Journal, 17*, 354-381.

Tsimhoni, O., Smith, D., & Green, P. (2004). Address entry while driving: Speech recognition versus a touch-screen keyboard. *Human Factors, 46*(4), 600-610.

Veltman, J. A., & Gaillard, A. W. K. (1998). Physiological workload reactions to increasing levels of task difficulty. *Ergonomics, 41*(5), 656–669.

Victor, T., Bärgman, J., Boda, C., Dozza, M., Engström, J., Flannagan, C., Lee, J.D. & Markkula, G. (2014). Analysis of Naturalistic Driving Study Data: Safer Glances, Driver Inattention, and Crash Risk. (SHRP 2 Safety Project S08A Prepublication Draft), Gothenburg, Sweden: Safer vehicle and traffic safety centre at Chalmers.

Wilson, G. F. (2002). Psychophysiological test methods and procedures. In S. G. Charlton & T. G. O'Brien (Eds.), *Handbook of human factors testing and evaluation* (pp. 127–156). Mahwah, NJ: Lawrence Erlbaum.

Appendix

Table 5. Means (and standard errors) by phone calling task and embedded vehicle system (Cheverolet MyLink or Volvo Sensus).

	Vehicle	Phone Easy (Manual)	Phone Hard (Manual)	Phone Easy (Voice)	Phone Hard (Voice)
Self-Reported Workload	Chevrolet	4.28 (0.4)	5.20 (0.4)	1.81 (0.3)	1.90 (0.3)
	Volvo	5.49 (0.4)	6.12 (0.4)	2.05 (0.2)	2.55 (0.3)
Task Completion Time	Chevrolet	29.18 (2.0)	23.30 (0.9)	20.48 (1.3)	22.74 (1.8)
	Volvo	31.43 (2.0)	34.36 (2.6)	34.48 (1.3)	41.87 (1.6)
% Change in Heart Rate	Chevrolet	2.54 (0.9)	1.12 (0.8)	2.46 (0.8)	3.75 (0.8)
	Volvo	2.07 (1.0)	2.10 (0.7)	2.47 (0.7)	0.97 (0.6)
% Change in SCL	Chevrolet	15.06 (3.8)	13.15 (3.0)	13.66 (3.3)	12.22 (3.2)
	Volvo	13.09 (3.0)	12.75 (2.9)	7.62 (2.8)	3.63 (2.7)
Mean Off-Road Glance Duration	Chevrolet	0.89 (0.0)	0.94 (0.0)	0.60 (0.0)	0.61 (0.0)
	Volvo	0.94 (0.0)	0.95 (0.0)	0.79 (0.0)	0.79 (0.0)
Mean Glance to Device Duration	Chevrolet	0.91 (0.0)	0.97 (0.0)	0.37 (0.0)	0.37 (0.1)
	Volvo	0.97 (0.0)	0.98 (0.0)	0.85 (0.0)	0.86 (0.0)
% of Off-Road Glances > 2.0s	Chevrolet	1.73 (0.6)	3.12 (0.8)	0.09 (0.1)	0.49 (0.5)
	Volvo	2.98 (1.0)	3.80 (1.0)	0.94 (0.4)	0.57 (0.2)
% of Glances to Device > 2.0s	Chevrolet	1.77 (0.6)	3.40 (0.9)	0.05 (0.1)	1.46 (1.5)
	Volvo	3.33 (1.2)	4.12 (1.1)	1.13 (0.6)	0.90 (0.4)
Total Off-Road Glance Time	Chevrolet	15.16 (1.2)	11.97 (0.7)	3.42 (0.5)	3.23 (0.4)
	Volvo	15.95 (1.1)	16.82 (1.4)	9.78 (0.7)	10.65 (0.9)
Total to Device Glance Time	Chevrolet	14.41 (1.2)	11.44 (0.6)	1.61 (0.4)	1.26 (0.2)
	Volvo	15.35 (1.1)	15.99 (1.4)	7.53 (0.6)	7.99 (0.8)
Number of Off-Road Glances	Chevrolet	16.74 (1.1)	12.82 (0.6)	5.26 (0.7)	5.05 (0.6)
	Volvo	16.96 (1.0)	17.70 (1.3)	12.44 (1.0)	13.46 (1.0)
Number of Glances to Device	Chevrolet	15.39 (1.0)	11.79 (0.5)	2.19 (0.4)	1.76 (0.3)
	Volvo	15.89 (1.0)	16.24 (1.2)	8.84 (0.8)	9.31 (0.8)
Speed (CAN - MPH)	Chevrolet	108.86 (2.0)	107.82 (2.4)	111.71 (1.4)	112.89 (1.5)
	Volvo	105.06 (1.2)	105.83 (1.6)	107.53 (1.1)	106.72 (1.3)
% Change in Speed (CAN)	Chevrolet	-3.77 (1.6)	-4.71 (1.9)	-1.19 (1.0)	-0.17 (1.1)
	Volvo	-3.94 (0.9)	-3.30 (1.3)	-1.70 (0.7)	-2.37 (1.1)
Speed (GPS - KPH)	Chevrolet	105.90 (2.0)	104.83 (2.3)	108.65 (1.4)	109.71 (1.5)
	Volvo	107.45 (1.3)	109.58 (1.0)	110.05 (1.1)	109.02 (1.4)
% Change Speed (GPS)	Chevrolet	-33.48 (5.6)	-41.58 (4.3)	-54.76 (4.3)	-53.90 (3.6)
	Volvo	-27.36 (9.4)	-36.24 (3.8)	-30.57 (4.2)	-22.80 (6.3)
Standard Deviation of Speed	Chevrolet	2.99 (0.3)	2.58 (0.2)	1.93 (0.2)	1.95 (0.2)
	Volvo	3.04 (0.5)	2.63 (0.2)	2.85 (0.2)	2.87 (0.2)
% Change in SD of Speed	Chevrolet	-36.31 (5.8)	-43.11 (4.7)	-57.46 (4.5)	-57.28 (3.6)
	Volvo	-30.49 (9.5)	-37.71 (4.3)	-33.73 (3.9)	-32.59 (4.4)
Major Wheel Reversals per minute	Chevrolet	25.65 (1.9)	26.79 (1.8)	26.37 (1.8)	23.22 (1.9)
	Volvo	4.78 (0.4)	4.10 (0.5)	3.74 (0.4)	4.05 (0.3)
% Change in Major Wheel Reversals	Chevrolet	23.26 (8.4)	27.58 (7.2)	28.63 (9.2)	16.80 (10.5)
	Volvo	51.41 (13.4)	38.11 (16.0)	31.31 (17.0)	37.84 (14.8)

Table 6. Summary of ANOVA-by-ranks on the phone tasks for variables of vehicle, modality, and vehicle x modality. *** = p < .001; ** = p < .01; * = p < .05; + = borderline effect (p < .10); NS = not significant.

Variable	Modality	System	Modality x System
Self-Reported Workload	***	NS	NS
Task Completion Time	NS	***	***
% Change in Heart Rate	NS	NS	+
% Change in SCL	+	NS	NS
Mean Off-Road Glance Duration	***	**	***
Mean Glance to Device Duration	***	***	***
% of Off-Road Glances > 2.0sec	***	NS	NS
% of Glances to Device > 2.0sec	***	NS	NS
Total Off-Road Glance Time	***	***	***
Total Glance to Device Time	***	***	***
Number of Off-Road Glances	***	***	***
Number of Glances to Device	***	***	***
Speed (CAN – MPH)	**	***	NS
Speed (GPS - KPH)	**	NS	NS
% Change in Speed (CAN)	**	NS	NS
%Change in Speed (GPS)	**	NS	NS
Standard Deviation of Speed	*	**	***
% Change in SD of Speed	*	***	***
Major Wheel Reversals	NS	***	NS
% Change in Major Wheel Reversals	NS	NS	NS

Table 7. Means (and standard errors) and results of Wilcoxon signed rank tests for variable measured during the destination address entry task periods. Change scores represent the percent (%) change from baseline just driving. Where presented, percent change values are likely to provide a more accurate representation of relative change for a particular variable as discussed in the body of the paper.

	Chevrolet	**Volvo**	**W**	**P-value**	
Self-Reported Workload	3.59 (0.44)	2.54 (0.28)	924.5	0.154	
Task Completion Time	66.68 (2.85)	80.60 (1.71)	408	< 0.001	*
% Change in Heart Rate	1.66 (0.87)	1.25 (0.67)	801	0.996	
% Change in SCL	11.59 (3.77)	3.29 (2.44)	811	0.172	
Mean Off-Road Glance Duration	0.74 (0.02)	0.82 (0.02)	562	0.022	*
Mean Glance to Device Duration	0.81 (0.04)	0.91 (0.03)	575	0.03	*
% of Off-Road Glances > 2.0sec	1.02 (0.29)	1.27 (0.36)	777.5	0.813	
% of Glances to Device > 2.0sec	1.48 (0.48)	1.83 (0.54)	747.5	0.569	
Total Off-Road Glance Time	14.28 (1.22)	22.56 (1.43)	367	< 0.001	*
Total Glance to Device Time	8.25 (0.89)	15.80 (1.16)	305	< 0.001	*
Number of Off-Road Glances	18.65 (1.52)	27.77 (1.75)	397	< 0.001	*
Number of to Device Glances	9.16 (0.98)	17.43 (1.25)	286.5	< 0.001	*
Speed (CAN - MPH)	113.68 (1.06)	108.26 (0.88)	1186	< 0.001	*
Speed (GPS - Knots)	59.65 (0.56)	59.85 (0.50)	797	0.981	
% Change in Speed (CAN)	0.57 (0.62)	-1.01 (0.54)	997	0.058	
% Change in Speed (GPS)	0.60 (0.62)	-0.98 (0.55)	990	0.068	
Standard Deviation of Speed	3.24 (0.18)	3.85 (0.25)	649	0.148	
% Change in SD of Speed	-29.53 (4.58)	-10.35 (5.46)	550	0.016	*
Major Wheel Reversals	21.78 (1.22)	2.88 (0.23)	1598	< .001	*
% Change in Major Wheel Reversals	7.34 (6.10)	-8.32 (6.82)	1003	0.051	

*p<.05

Accident Analysis and Prevention 52 (2013) 125–132

Contents lists available at SciVerse ScienceDirect

Accident Analysis and Prevention

journal homepage: www.elsevier.com/locate/aap

Impact of age and cognitive demand on lane choice and changing under actual highway conditions

Bryan Reimer[a,*], Birsen Donmez[a,c], Martin Lavallière[a,b], Bruce Mehler[a], Joseph F. Coughlin[a], Normand Teasdale[b]

[a] *MIT AgeLab & New England University Transportation Center, 77 Massachusetts Avenue, E40-279 Cambridge, MA 02139, USA*
[b] *Université Laval, Faculté de Médecine, & Vieillissement, Centre de recherche FRSQ du Centre hospitalier affilié universitaire de Québec, 2300, rue de la Terrasse, Québec, Québec, Canada*
[c] *University of Toronto, Department of Mechanical and Industrial Engineering, 5 King's College Road, Toronto, ON M5S 3G8, Canada*

ARTICLE INFO

Article history:
Received 18 January 2011
Received in revised form
24 November 2012
Accepted 7 December 2012

Keywords:
Lane change behavior
Lane choice
Distraction
Cognitive workload
Passing
Aging

ABSTRACT

Previous research suggests that drivers change lanes less frequently during periods of heightened cognitive load. However, lane changing behavior of different age groups under varying levels of cognitive demand is not well understood. The majority of studies which have evaluated lane changing behavior under cognitive workload have been conducted in driving simulators. Consequently, it is unclear if the patterns observed in these simulation studies carry over to actual driving. This paper evaluates data from an on-road study to determine the effects of age and cognitive demand on lane choice and lane changing behavior. Three age groups (20–29, 40–49, and 60–69) were monitored in an instrumented vehicle. The 40's age group had 147% higher odds of exhibiting a lane change than the 60's group. In addition, drivers in their 60's were less likely to drive on the leftmost lane compared to drivers in their 20's and 40's. These results could be interpreted as evidence that older adults adopt a more conservative driving style as reflected in being less likely to choose the leftmost lane than the younger groups and less likely to change lanes than drivers in their 40's. Regardless of demand level, cognitive workload reduced the frequency of lane changes for all age groups. This suggests that in general drivers of all ages attempt to regulate their behavior in a risk reducing direction when under added cognitive demand. The extent to which such self-regulation fully compensates for the impact of added cognitive demand remains an open question.

1. Introduction

Driving is a complex skill that can be considered as a combination of different functional and operational activities involving low level control of the vehicle guided by maneuvers and strategic decisions (Michon, 1995). Lane changing is a driving maneuver frequently associated with accidents (Pande and Abdel-Aty, 2006) and requires engagement of a coordinated combination of sensory/perceptual, cognitive processing, and manipulative actions. Humans are generally considered to have finite information processing resources (Wickens, 1984; Wickens & McCarley, 2008), and situations that make multiple calls on these resources, particularly those that require divided attention, may tax capacity to the point that performance and safety margins suffer. Possible strategies for coping with increased demand might include limiting overall workload by reducing the frequency of optional maneuvers, such as non-critical lane changes, or by actions such as slowing driving speed.

In a naturalistic study of 16 commuters using either interstate or state highways in southwestern Virginia, Olsen et al. (2002) found that lane changes were most often initiated due to a slow lead vehicle and occurred more frequently on the interstate. In a driving simulator study, Bar-Gera and Shinar (2005) observed that their subjects frequently passed lead vehicles that were faster than their own average speed. The authors suggested that their subjects chose to perform the passing maneuver to minimize the workload associated with following a lead vehicle. Thus, the speed differential between a lead vehicle and the driver's own vehicle is not the only factor that guides lane changing decisions. Aggressiveness, sensation seeking, and competitiveness have also been suggested to affect lane change behavior (Bar-Gera & Shinar, 2005; Matthews et al., 1998).

The level of cognitive demand is another factor that can influence lane changing decisions and requires particular attention

* Corresponding author. Tel.: +1 617 452 2177.
E-mail addresses: reimer@mit.edu (B. Reimer), donmez@mie.utoronto.ca (B. Donmez), martin.lavalliere.1@ulaval.ca (M. Lavallière), bmehler@mit.edu (B. Mehler), coughlin@mit.edu (J.F. Coughlin), Normand.Teasdale@kin.ulaval.ca (N. Teasdale).

0001-4575/$ – see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.aap.2012.12.008

as technological developments create additional possibilities for drivers to engage in cognitively demanding activities. In driving simulators, participation in a cognitive secondary task has been shown to interfere with the frequency of lane changes (Beede & Kass, 2006; Cooper et al. (2009)) and to affect the degree to which drivers monitor surrounding traffic conditions (Zhou et al., 2009). In Cooper et al. (2009), the effect of a hands-free cell phone conversation on lane changing was investigated in three levels of traffic density. The results show that when drivers conversed on a cell phone, they made fewer lane changes, were more likely to remain behind a slow vehicle, and had a lower overall mean speed. Beede and Kass (2006) reported similar results in which drivers changed lanes most frequently when not engaged in any secondary activities. The changes in driving behavior observed by Cooper et al. (2009) and Beede and Kass (2006) may be viewed as compensatory actions taken by the driver to reduce the workload associated with the driving task and diverting extra capacity to other non-driving activities. Alternatively, these changes in behavior might be interpreted as an effect of saturation in cognitive capacity, which results in reduced attention to elective driving activities such as changing lanes.

Cantin et al. (2009) assessed mental workload under various driving conditions in a simulator study by measuring reaction times to periodic auditory probes and found that reaction times during the execution of a lane change is slower than during periods without lane changes. The relative increase in workload, as evidenced by increased response delay was greater for older adults. As drivers age, they self-report that they are less likely to pass another vehicle (Boyle et al., 1998). In a driving simulator, Farah and Toledo (2010) found that drivers under age 25 tend to accept smaller passing gaps and are more likely to pass a lead vehicle compared to drivers over age 25. Under actual driving conditions, older drivers are known to self-regulate workload, e.g., drive slower, travel during less congested periods, and avoid distracting technologies (D'Ambrosio et al., 2008; Langford & Koppel, 2006). Less is known, however, about the lane changing behavior of different age groups under varying levels of cognitive demand. Finally, the degree to which lane changing behavior with cognitive workload observed in previous driving simulator studies carries over to field driving is not well established.

This paper expands upon recently published work on physiological reactivity and changes in visual behavior in response to graded levels of cognitive demand across different age groups (Mehler et al., 2012; Reimer et al., 2012) by examining lane changing behavior in the same dataset. The data were captured during an extended period of driving during non-rush hour, daytime periods on a multi-lane interstate during which participants were free to maintain or adjust their lane positioning at will. The sample was limited to cases that were classified as being largely unaffected by traffic flow (e.g., cases with extended periods of stop and go traffic and cases impacted by complete gridlock were excluded) or adverse weather conditions. The dataset therefore contains behaviors that occurred during periods of continuous traffic flow where the decision to make lane changes was effectively at the drivers' discretion. These data provided an opportunity to examine the frequency of lane changing and lane selection under both single task driving and under conditions of objectively defined levels of secondary cognitive demand. The age groups studied allow for the characterization of lane changing behavior and lane choice across the lifespan. This work also provides an opportunity to examine additional aspects of the extent to which drivers of different ages do or do not compensate for the added demand of secondary cognitive workload during highway driving as well as a validation point for observations from controlled simulation studies.

2. Methods

2.1. Participants

One-hundred and sixty five participants initially reported for the study. Participants were self-reported experienced drivers, driving more than three times a week and having held a valid driver's license for over 3 years. Inclusion criteria required participants to be free of police reported accidents for the past year, making the group limited to a potentially safer set of drivers than would be observed in a broader community sample. The group was considered to be relatively healthier than a community sample as individuals were excluded if they reported a recent hospitalization, specific health conditions (such as a positive history for heart attack, angina, coronary heart disease, stroke, a pacemaker, or uncontrolled diabetes), recent use of medications that cause drowsiness or suggest safety concerns (e.g., anti-psychotic or anti-convulsant medications), or medications that impact heart rate (e.g., beta blockers). Additionally, participants with scores below 26 on the Mini-Mental State Examination (Folstein et al., 1975), an indication of cognitive impairment, were excluded. Compensation of $60 was provided for the 3-h experiment.

Cases were excluded from the final dataset as follows: failure to meet requirements upon eligibility review (7), inability to perform the secondary tasks to criterion during training (8) [4 each in the 40's and 60's age categories], pilot runs (first 4), sleepy while driving (3), heavy traffic (10), adverse weather during a portion of the drive (8), protocol errors (5), equipment failures (7), poor video quality (2), withdrew from study after learning requirements (2), and extra cases dropped after design cells (age and gender) were filled (3). The sample considered here consisted of the remaining 106 individuals and was balanced by gender and across three age groups: 20–29 ($n = 36$), 40–49 ($n = 35$), and 60–69 ($n = 35$). The average age by group was 24.6 (SD: 2.7), 44.4 (SD: 3.0), and 63.3 (SD: 3.1). Male and female participants did not differ significantly by age within each group ($F(1,34) = .86$, $p = .36$; $F(1,33) = .83$, $p = .37$; $F(1,33) = .22$, $p = .64$).

2.2. Apparatus and secondary task

Participants drove an instrumented mid-sized sports utility vehicle (Volvo XC 90) equipped for time synchronized data collection. Data presented here were recorded from the vehicle's CAN bus, a microphone mounted inside the vehicle, and from a camera mounted near the center of the vehicle facing forward. Three levels of an auditory presentation – verbal response, delayed digit recall task (n-back) were employed to increase drivers' workload. Each level consisted of four 30-s blocks during which 10 single digit numbers (0–9) were presented in random order at a spacing of 2.25 s. At the lowest level of demand (0-back), drivers were to repeat each digit as it was presented. At the moderate level of demand (1-back), drivers were to respond to each new presentation by recalling and saying out loud the previous number in the presentation sequence. At the highest level of demand (2-back), the number two places back in the sequence was to be repeated. The form of this task was identical to earlier studies conducted in our laboratory (Mehler et al., 2009; Reimer, 2009; Reimer & Mehler, 2011) and was developed based on recommendations by Zeitlin (1993) for secondary cognitive tasks for use in driving contexts. This form of n-back task holds the amount of auditory demand constant across levels, verbal resources required for responding are essentially the same, and only the memory demands change – increasing in a systematic manner with each level of the task. For additional details on the development of this task, training materials, and the complete stimulus sets see Mehler et al. (2011).

2.3. Procedure

Participants signed an approved informed consent and completed a questionnaire covering driving and health history. They were then extensively trained on the secondary task prior to entering the vehicle with a minimum of $n+1$ practice blocks per task level (e.g., 3 practice blocks for the 2-back). Additional repetitions of the instructions and practice trials were presented for each demand level until participants demonstrated a minimum proficiency of 7 correct responses on the 0 and 1-back (out of 10 and 9 items respectively) and of at least 4 (out of 8) on the 2-back. Participants who were unable to meet the criteria for the 2-back within nine practice blocks were excluded (8 individuals). Upon being seated in the vehicle, an audio recording reviewed the secondary task and presented an additional $n+3$ practice blocks for each demand level.

A research associate, seated in the back of the car, operated the data collection equipment, provided driving directions, and monitored participants to ensure that they had adequate control of the vehicle at all times. Approximately 30 min of driving were provided for habituation prior to the study period. The study period consisted of an initial 6 min of single task driving followed by the three levels of the n-back task. Each task was presented over a 2-min interval (four 30-s blocks) and each task was followed by 2 min of single task driving. The presentation order for the three levels of the task was counterbalanced across the sample.

The experiment was conducted on Interstate 93 starting in the vicinity of the intersection with I495. Participants were traveling north from Boston, Massachusetts to the area generally just south of Manchester, New Hampshire. The posted speed limit was 104.6 km/h (65 mph). When first entering the highway in Boston, Massachusetts, participants were prompted: "We are going to be driving north on 93 for approximately 40 min. You can continue driving in this lane or move into another lane so that you are comfortable with the traffic flow." Participants were thus allowed to select a lane of travel and pass other vehicles at will. Instances of weather or other traffic conditions that impeded normal traffic flow or otherwise constrained a driver's ability to change lanes were recorded by the research assistant in the experimental log and reviewed by a senior staff member. As noted previously, a total of 18 cases were dropped from consideration due to weather or traffic density constraints on normal traffic flow.

The distribution of the number of travel lanes across the sample was not uniform due to the nature of variations in traffic and driver speed that affected the start of the data assessment period for individual drivers. Excluding exits, on-ramps and other road transitions, the roadway began with four lanes of travel, decreased to three and, eventually in Southern New Hampshire, to two. Then, near the end of the portion of highway used in the study, the number of lanes increased back to three. Twelve participants began the experimental period while in a section of four lane highway, 90 in three lanes and 4 in two lanes. Twenty five participants completed the experimental segment while still on a portion of the highway with two travel lanes.

2.4. Data coding and data periods used in analysis

Recorded audio was used to assess participants' accuracy in responding to the n-back tasks. Task performance was scored as a percentage of the number of correct responses out of the total number of expected responses. Lane change and lane choice data were independently extracted through a manual analysis of video recordings by two research associates. Discrepancies were than reviewed and reconciled by the first author. No differentiation between critical and non-critical lane changes was made. The procedure for classification of lane changes was analogous to Olsen et al. (2002) and consistent with Cooper et al. (2009) where the onset of each lane change was classified as the point when the vehicle was observed to be first moving in a lateral direction toward the destination lane. The completion of each lane change was recorded as the point where the vehicle was fully centered in the destination lane. Time spent in the leftmost lane was computed as the time from the lane crossing entering the left lane to the lane crossing exiting the left lane. A lane crossing was classified as when the middle of the car crossed the lane marker. Only lane changes that resulted in the centerline of the vehicle crossing over the dividing line were considered; partial motions toward an adjacent lane were not coded.

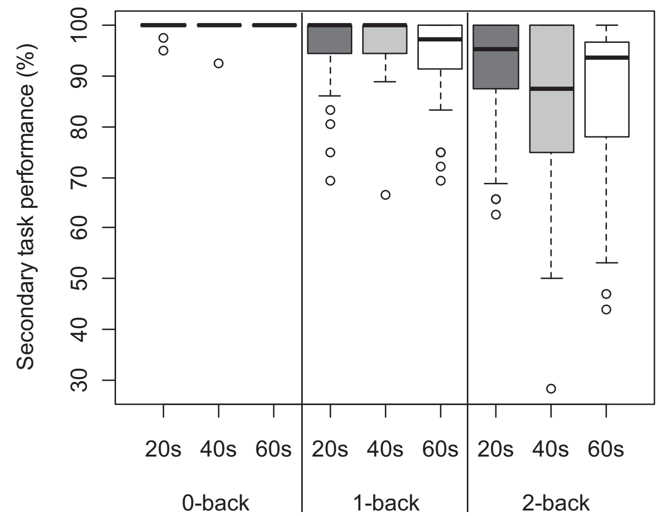


Fig. 1. Secondary task performance (%) for each age and demand level (0, 1, and 2-back task)*.

* Figures 1, 3, and 4 are boxplots which represent the five-number summary (minimum, first quartile, median, third quartile, and maximum) as well as potential outliers as indicated with circles.

Three data periods were analyzed: pre-task, n-back (cognitive demand tasks), and recovery. The pre-task period consisted of 6 min of single task driving prior to the initiation of the first n-back task. The n-back period consisted of the 6 min of dual task activity corresponding to the aggregate of the three separate 2-min-long secondary tasks. The 6 min of data for the recovery period were drawn from the 2 min of single task driving that followed each of the three dual task periods. Also considered was an analysis of each of the three task demand levels (0, 1, and 2-back). Since each task was 2 min long, a 2-min reference "baseline" period was used for comparison. Consistent with Mehler et al. (2012) and Reimer et al. (2012), the 2-min baseline period was selected as minutes 3.5 to 5.5 of the pre-task period. For the purpose of classifying frequencies, a lane change was assigned to the period in which it was initiated. Gender was initially included in all statistical analysis, but was later dropped from the final models, as it was not a significant predictor in any of the models.

3. Results

3.1. Secondary task performance

Fig. 1 presents the secondary task performance for each age group by the level of cognitive task difficulty. The rate of correct responses was analyzed with a Poisson model due to the high level of non-normality in the data. There were 40, 36, and 32 stimuli that required responses in the 0-back, 1-back, and 2-back conditions, respectively. The logarithm of the number of stimuli was used as an offset variable in the model. Because the data consisted of repeated measures, generalized estimating equations (GEE) were used for estimation. The model was fitted using PROC GENMOD in SAS 9.1, with the specifications of log link function and Poisson distribution. Significant effects were observed for cognitive task difficulty ($\chi^2(2) = 79.03$, $p < .0001$) and its interaction with age ($\chi^2(4) = 10.18$, $p = .04$). Follow-up contrasts revealed that, regardless of age, increasing demand resulted in degraded secondary task performance (0-back vs. 1-back: $\chi^2(1) = 37.0$, $p < .0001$; 0-back vs. 2-back: $\chi^2(1) = 64.5$, $p < .0001$; 1-back vs. 2-back: $\chi^2(1) = 30.4$, $p < .0001$). This effect suggests that as in Mehler et al. (2012), Reimer (2009), and Reimer and Mehler (2011), error rates increased with higher levels of cognitive task difficulty. The differences between age groups depended on the cognitive task demand. Under the highest demand condition (2-back), the 20's age group responded correctly to a higher percentage of stimuli than both the 40's ($\chi^2(1) = 4.0$, $p = .045$) and 60's ($\chi^2(1) = 3.11$, $p = .078$) age groups, although the comparison to the 60's group was only marginally significant. There were no significant differences across the age groups for the 0-back or 1-back conditions ($p > .1$).

It is appropriate to note again that 4 potential participants from each of the 40's and the 60's groups were excluded from data collection due to difficulties with the high demand 2-back task during training. We only wanted to consider drivers who were able to engage with the task and thus actually experience increased cognitive load. Consequently, the task performance levels for the two older age groups may well have been somewhat higher than what would be observed in unscreened samples.

3.2. Number of lane changes

A negative binomial model was developed to compare the number of lane changes across different age groups for three study periods: the 6-min period prior to the admission of the n-back tasks (pre-task), the aggregate of the three 2-min-long n-back tasks (n-back), and the combination of the three 2-min intervals of single task driving following each n-back task (recovery). Each of these three periods was 6-min in duration. The model was fitted using PROC GENMOD in SAS 9.1, with the specifications of a log link function and negative binomial distribution. Repeated measures were accounted for by using GEE. Fig. 2 shows histograms for the number of lane changes across different age groups and study periods.

Both the main and interaction effects were statistically significant ($p < .05$). For the 20's age group, the lowest number of lane changes occurred during the n-back period, followed by the pre-task period, and the highest number of lane changes was observed during the recovery period. For this age group, the expected number of lane changes in the recovery period was estimated to be 2.39 times the expected number of lane changes in the n-back period (95% CI: 1.70, 3.35, $\chi^2(1) = 25.25$, $p < .0001$) and 1.45 times the expected number of lane changes in the pre-task period (95% CI: 1.05, 2.00, $\chi^2(1) = 5.02$, $p = .03$). The expected number of lane changes in the pre-task period was 1.66 times that in the n-back period (95% CI: 1.07, 2.58, $\chi^2(1) = 5.03$, $p = .03$).

For the 40's age group, the only significant difference was between the recovery and n-back periods with a multiplicative increase of 1.34 in the expected number of lane changes in the recovery period (95% CI: 1.003, 1.80, $\chi^2(1) = 3.93$, $p = .048$).

For the 60's age group, n-back period resulted in lower number of lane changes compared to both the pre-task and recovery periods. Compared to the n-back period, a multiplicative increase of 2.17 and 2.38 were observed in the expected number of lane changes in the pre-task and recovery periods, respectively (95% CI: 1.35, 3.50, $\chi^2(1) = 10.15$, $p = .001$; 95% CI: 1.47, 3.84, $\chi^2(1) = 12.59$, $p = .0004$).

Comparisons of different periods across age groups did not reveal major significant findings, likely due to the between subject nature of these comparisons. The only statistically significant finding was that, during the n-back period, the 40's group had a higher number of lane changes compared to the 60's group: a multiplicative increase of 2.31 in the expected number of lane changes (95% CI: 1.24, 4.31, $\chi^2(1) = 6.94$, $p = .008$).

Table 1
Number of lane changes by age and demand level (the cells represent the number of drivers who performed a given number of lane changes).

Age	Demand level	Number of lane changes					
		0	1	2	3	4	5
20's ($n = 36$)	Baseline	21	7	5	2	1	0
	0-Back	25	7	2	2	0	0
	1-Back	25	9	1	1	0	0
	2-Back	26	7	3	0	0	0
	Total	97	30	11	5	1	0
40's ($n = 35$)	Baseline	18	8	8	0	1	0
	0-Back	19	11	4	0	0	1
	1-Back	16	8	10	1	0	0
	2-Back	26	6	3	0	0	0
	Total	79	33	25	1	1	1
60's ($n = 35$)	Baseline	23	7	4	1	0	0
	0-Back	28	5	1	1	0	0
	1-Back	28	4	3	0	0	0
	2-Back	27	7	1	0	0	0
	Total	106	23	9	2	0	0

A finer break-down of the n-back task period is provided in Table 1. These data are based on the 2-min windows represented by the three separate cognitive difficulty levels (0, 1 and 2-back) and where "baseline" represents 2-min prior to the first n-back task. Given the short intervals, there were only a few cells with more than one lane change. Thus, for analysis purposes, we grouped the response variable into two categories: 0 and $\geq$1 lane changes. A logistic regression model was built to predict the likelihood of making at least one lane change for different age groups and demand levels. Repeated measures were accounted for using GEE. The model was fitted using PROC GENMOD in SAS 9.1, with the specifications of logit link function and binomial distribution.

Wald statistics revealed that age ($\chi^2(2) = 7.81$, $p = .02$) and demand level ($\chi^2(3) = 8.24$, $p = .04$) were statistically significant. The interaction of age with demand level was not significant ($p > .05$). The 40's group exhibited 147% higher odds of making lane changes compared to the 60's group (95% CI: 29%, 372%, $\chi^2(1) = 7.44$, $p = .006$). No other differences were found between age groups. These findings are to some extent in line with the results of the negative binomial model reported above for the different study periods. As for demand level, drivers had 112% higher odds of making lane changes without the cognitive task compared to with the 2-back task (95% CI: 24%, 262%, $\chi^2(1) = 7.63$, $p = .006$).

3.3. Mean speed

A repeated measures ANOVA was conducted to compare mean speed across the three age groups for the three study periods previously discussed: pre-task, n-back, and recovery (Fig. 3). The main effects of age ($F(2,102) = 5.53$, $p = .005$) and study period ($F(2,204) = 16.79$, $p = <.0001$) and their interaction ($F(4,204) = 3.91$, $p = .004$) were all significant. In general, participants in the 20's age group drove significantly faster than those in the 60's group ($t(102) = 3.06$, $p = .003$). The 40's group also drove significantly faster than the 60's group, but only during the n-back task ($t(169) = 3.08$, $p = .003$) and recovery ($t(169) = 3.39$, $p = .001$) periods and not during pre-task. The 20's and 60's groups drove significantly faster prior to the n-back task period than they drove during n-back task and recovery periods (20's pre-task vs. 20's n-back:

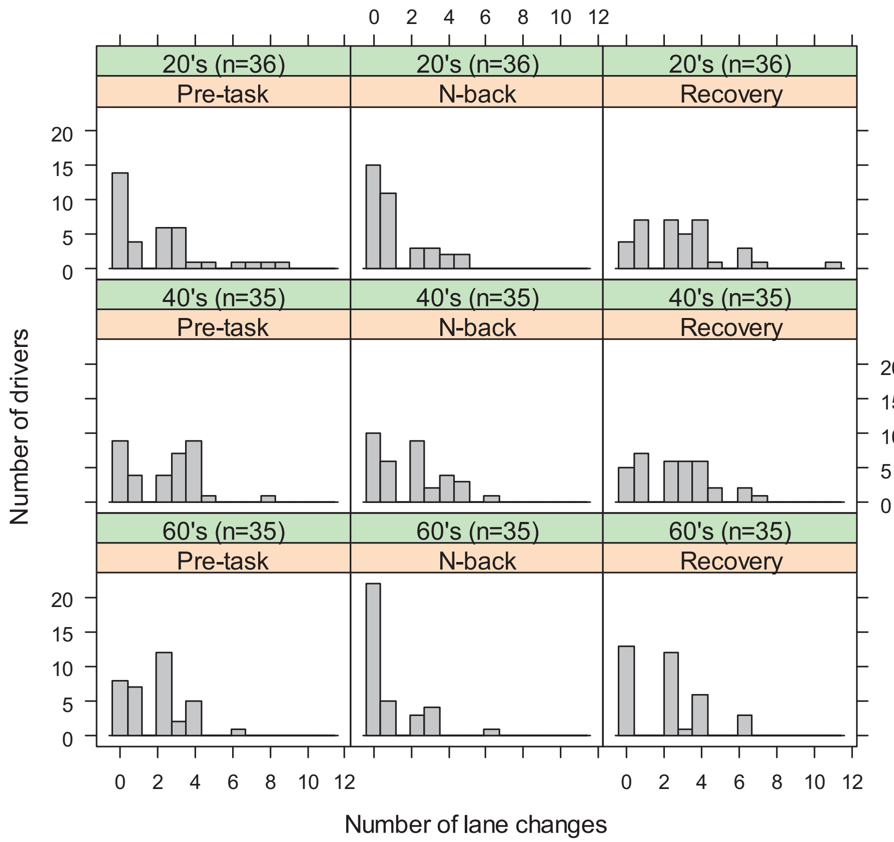


Fig. 2. Histograms for number of lane changes across age groups and study periods.

$t(204) = 4.05$, $p < .0001$; 20's pre-task vs. 20's recovery: $t(204) = 3.77$, $p = .0002$; 60's pre-task vs. 60's n-back: $t(204) = 4.72$, $p < .0001$; 60's pre-task vs. 60's recovery: $t(204) = 4.39$, $p < .0001$). Thus, the results suggest that in general the 40's group's average speed profile stayed fairly constant across the three study periods, whereas the 20's and 60's groups reduced their speeds when presented with the n-back task and maintained these slower speeds during the recovery periods. These results are based on aggregate speed profiles for 6-min study periods. A more detailed analysis on speed maintenance across the three demand levels follows.

Fig. 4 presents the mean speed for each category of age and demand level. The repeated measures ANOVA yielded significant main effects of age ($F(2,103) = 5.05$, $p = .008$) and demand level ($F(3,309) = 8.79$, $p < .0001$). The interaction between age and demand level was not significant ($p > .05$). Compared to the 20's and 40's age groups, drivers in their 60's drove significantly slower (105.4 and 105.3 vs. 101.7 km/h, respectively) (60's vs. 20's: $t(103) = -2.78$, $p = .006$; 60's vs. 40's: $t(103) = -2.73$, p=.008). During single task driving (baseline), drivers of all ages drove faster than during all three demand level periods (baseline: 106.1 km/h, 0-back: 103.9 km/h, 1-back: 104 km/h, and 2-back: 102.3 km/h)(baseline vs. 0-back: $t(309) = 2.87$, $p = .004$; baseline vs. 1-back: $t(309) = 2.76$, $p = .006$; baseline vs. 2-back: $t(309) = 5.12$, $p < .0001$). The 0- and 1-back tasks both resulted in faster average speeds than the 2-back task (0-back vs. 2-back: $t(309) = 2.26$, $p = .02$; 1-back vs. 2-back: $t(309) = 2.36$, $p = .02$).

3.4. Lane choice

Table 2 presents the number of drivers who did or did not at all drive on the leftmost lane (high speed lane) across the three study periods. A logistic regression model was built to predict the

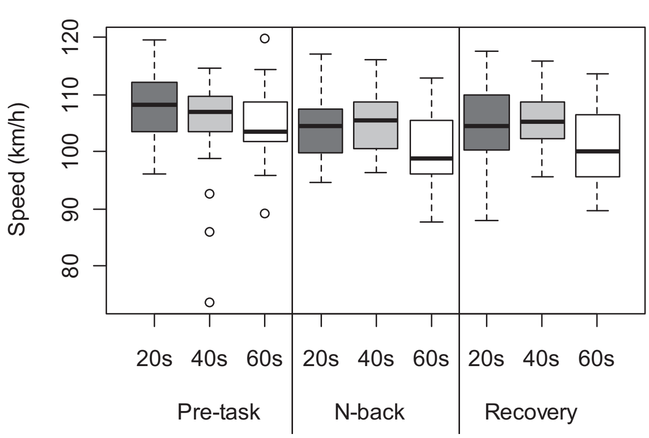


Fig. 3. Mean speed across different age groups and study periods.

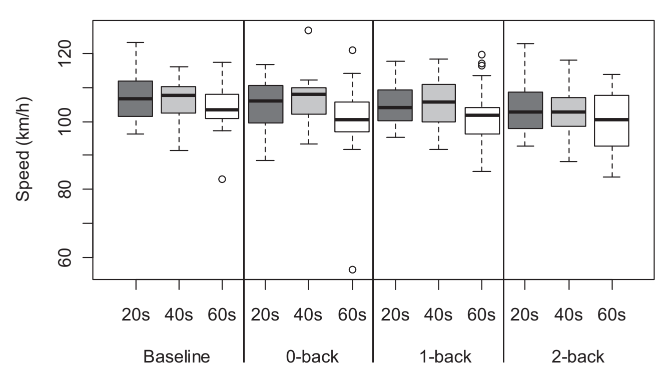


Fig. 4. Mean speed for each age and demand level (baseline, 0, 1, and 2-back task).

Table 2
Percentage of drivers who drove in the leftmost lane across age group and study period.

Age	Pre-task	n-Back	Recovery	Row average
20's ($n=36$)	66.7	80.6	94.4	80.6
40's ($n=35$)	82.9	80.0	92.3	85.7
60's ($n=35$)	51.4	40.0	62.9	51.4

Table 3
Percentage of drivers who drove on the leftmost lane across age group and demand level.

Age	Baseline	0-Back	1-Back	2-Back	n-Back (mean)
20's ($n=36$)	58.3	52.8	58.3	50.0	53.7
40's ($n=35$)	60.0	57.1	65.7	42.9	55.2
60's ($n=35$)	28.6	25.7	22.9	31.4	26.7

likelihood of driving on the leftmost lane. Repeated measures were accounted for using GEE. The model was fitted using PROC GENMOD in SAS 9.1, with the specifications of logit link function and binomial distribution. The 60's group had lower odds of driving on the leftmost lane than both the 20's and the 40's group ($\chi^2(2)=21.19$, $p<.0001$). Compared to the 60's group, the 40's group was estimated to have a 544% and the 20's group was estimated to have a 390% higher odds of driving on the leftmost lane. The highest odds of driving on the leftmost lane was observed during the recovery period ($\chi^2(2)=12.99$, $p=.002$). The odds of driving on the leftmost lane were 259% and 250% higher in the recovery period compared to the pre-task and n-back periods, respectively. We also analyzed time spent on the leftmost lane for non-zero observations. There were no statistically significant findings for this variable ($p>.05$).

Table 3 presents the number of drivers who did or did not drive in the leftmost lane across different task demand levels. Another logistic regression model was built to analyze these data. Drivers in the 60's age group had lower odds of driving in the leftmost lane than both the 20's and the 40's groups ($\chi^2(2)=11.57$, $p=.003$). Compared to the 60's group, the 40's group was estimated to have a 251% and the 20's group a 229% higher odds of driving in the leftmost lane regardless of task demand. The analysis of time spent on the leftmost lane for non-zero observations did not reveal any statistically significant findings ($p>.05$).

4. Discussion

An analysis of data collected from an on-road experiment was conducted to examine lane change and lane choice behavior of three different age groups (20's, 40's and 60's) under varying levels of cognitive load. The findings reveal that both age and demand level were associated with lane choice and lane change behaviors. Compared to periods of single task driving, fewer lane changes were observed under secondary cognitive task load. A similar effect was observed by Cooper et al. (2009) in a simulator experiment. Given that our study was an on-road assessment, these results provide additional ecological validity to this finding. An age effect was observed in the significantly lower likelihood of making a lane change in the 60's age group compared to drivers in their 40's. This result suggests that, as expected, older adults in this study adopted a generally more conservative driving style than middle-aged adults. Contrary to expectations, the 20's group did not change lanes more frequently than the oldest group. This finding may in part relate to the sample, drawn from an urban population where younger drivers have generally lower and more variable levels of driving experience. It is plausible that individuals with less experience are not as willing to accept the risk associated with lane changes. In contrast, drivers in the 40's age group have more driving experience and may be more broadly willing to accept the risks associated with passing. An alternate explanation may be that the younger age group is less willing to engage in riskier behaviors such as changing lanes while driving an unfamiliar instrumented research vehicle in the presence of a research associate.

Speed selection also was influenced by age and demand level. The 60's group drove at a significantly slower mean speed than drivers in their 20's and 40's. This finding is in agreement with other on-road and simulator studies where older drivers maintained slower speeds than younger drivers (Hakamies-Blomqvist et al., 1999; Planek & Fowler, 1971; Szlyk et al., 1995). Moreover, older drivers in our study responded to the introduction of cognitive task load with the same compensatory strategy as younger drivers, that is, they reduced their speed by a similar amount. During the baseline, drivers travelled at greater speeds than they did when they had to perform each of the cognitive secondary tasks. These results are consistent with simulation studies by Cooper et al. (2009) and Horberry et al. (2006), which showed that drivers reduced their speed while conducting a conversation similar to a hands-free cell phone task.

A major concern around added cognitive demand during driving is that drivers' gaze concentrates around the road center (Harbluk et al., 2007; Reimer et al., 2012; Sodhi et al., 2002) and drivers are not always aware of the extent of the resulting attentional attenuation or any associated performance decrements (Horrey et al., 2008). Thus, they may be willing to engage in potentially risky distracting activities while driving, resulting in an overall reduction in their capacity to respond to various emerging demands in the driving environment. Concern over this pattern of behavior is one reason behind the call for the development of detection systems that provide information to drivers when they are distracted so that they can modify their behavior appropriately (Coughlin et al., 2011; Donmez et al., 2007, 2008). Donmez et al., 2010 found that young drivers who exhibit the riskiest distraction behavior among their peers benefit most from distraction-related feedback. This finding provides more evidence about drivers' lack of risk awareness when engaged in distracting activities and the potential benefits of providing feedback to guide appropriate behavior.

To our knowledge, this is the first time that driving in the leftmost lane as a function of age and cognitive load has been evaluated. The results show that the 60's age group was more likely to not drive on the leftmost lane at all compared to the 20's and 40's cohorts. This pattern remained consistent across single task driving and with the addition of the secondary cognitive task load. Self-regulation could explain why participants in the 60's age group differ from their younger counterparts in their utilization of the leftmost lane. Because the attentional and control demands of travel in this lane are generally greater due to a higher travel speed, avoiding the leftmost lane may be one strategy for reducing overall demand. Donorfio et al. (2009) used a survey to investigate self-regulation in older drivers. Older drivers described driving as a way to remain connected to society, and self-regulation represents one method they use to cope with changing capabilities due to declining health and cognitive abilities. In order to self-regulate appropriately, however, drivers need to be aware of their limitations and their capacity to find an effective balance. Unfortunately, older drivers are not always good at self-evaluation of their driving performance (Holland and Rabbitt, 1992, 1994), and older drivers have been found to err both on the side of driving beyond their capabilities and of sometimes curtailing their driving behavior prematurely (D'Ambrosio et al., 2008).

5. Conclusion

As noted above, this is, to our knowledge, the first attempt to assess if older drivers differ in their lane changing and lane choice behaviors compared to younger drivers in an on-road study employing a relatively large sample size. We also investigated the effects of varying levels of secondary task cognitive load on these behaviors. During periods of heightened cognitive load, lane change frequency was found to decrease across all age groups, and drivers in their 60's were found to make fewer lane changes than those in their 40's. In general, older drivers were less likely to make a lane change than drivers in their 40's and were less likely to drive on the left hand travel lane compared to drivers in their 20's and 40's. More research is required to determine if the observed reduction in lane changes represents a conservative, compensatory approach to driving under dual task load or simply results from a basic saturation of the drivers' cognitive capacities that limits engagement in other activities. In the latter case, the act of attending to the secondary task may by default reduce attention to the driving task, resulting in less active lane changing behavior and reduced pressure on the throttle resulting in speed reduction. In either case, a reduced frequency of higher risk behaviors, such as lane changing and higher speed, during periods of added cognitive demand may be one of many factors that explains why accidents are not increasing while drivers are being confronted with more visual, manipulative and cognitive distractions. Nonetheless, more research is needed to quantify the risks associated with non-driving related in-vehicle activities.

6. Limitations

The results presented here are limited to the extent that they focus on one type of cognitive demand and a single driving environment. The changing number of travel lanes across the drive introduced some forced lane changes. While forced lane changes were not included in the analysis, the driving environment may have introduced variability that impacted the results in subtle ways. Follow-on work considering a highway segment with a fixed number of lanes is warranted. The age effects observed in this study may be understated in that individuals were excluded due to reported health conditions, infrequent driving, recent police reported accidents and an inability to perform the n-back task to criterion during training, most likely resulting in a somewhat more frequent driving and higher functioning sample than would have been obtained without exclusions. Finally, the effect of driving a highly instrumented vehicle with an observer on lane changing and other driving behaviors is unknown. It might be that some drivers were less likely to make lane changes compared to what they would do alone in their own vehicles. In spite of these limitations, these results provide important external validity to earlier simulation findings.

Acknowledgements

We gratefully acknowledge the support of the US Department of Transportation's Region I New England University Transportation Center at MIT and The Santos Family Foundation during the analysis of these data and writing of this paper, and Ford Motor Company for the development of the original dataset and the instrumented vehicle. In addition, Martin Lavallière was supported by a doctoral scholarship and a FRSQ-RQRV International Internship Grant from the Fonds de la Recherche en Santé du Québec. An earlier version of this manuscript appeared in the *Proceedings of the Transportation Research Board 90*th *Annual Meeting* (Lavallière et al., 2010).

References

Bar-Gera, H., Shinar, D., 2005. The tendency of drivers to pass other vehicles. Transportation Research Part F: Traffic Psychology and Behaviour 8 (6), 429–439.

Beede, K.E., Kass, S.J., 2006. Engrossed in conversation: the impact of cell phones on simulated driving performance. Accident Analysis and Prevention 38 (2), 415–421.

Boyle, J., Dienstfrey, S., Sothoron, A., 1998. National Survey of Speeding and Other Unsafe Driving Actions. National Highway Traffic Safety Administration, Washington DC.

Cantin, V., Lavallière, M., Simoneau, M., Teasdale, N., 2009. Mental workload when driving in a simulator: effects of age and driving complexity. Accident Analysis and Prevention 41 (4), 763–771.

Cooper, J.M., Vladisavljevic, I., Medeiros-Ward, N., Martin, P.T., Strayer, D.L., 2009. An investigation of driver distraction near the tipping point of traffic flow stability. Human Factors 51 (2), 261–268.

Coughlin, J.F., Reimer, B., Mehler, B., 2011. Monitoring, managing, and motivating driver safety and well-being. IEEE Pervasive Computing 10 (3), 14–21.

D'Ambrosio, L.A., Donorfio, L.K.M., Coughlin, J.F., Mohyde, M., Meyer, J., 2008. Gender differences in self-regulation patterns and attitudes toward driving among older adults. Journal of Women and Aging 20 (3–4), 265–282.

Donmez, B., Boyle, L., Lee, J.D., 2008. Mitigating driver distraction with retrospective and concurrent feedback. Accident Analysis and Prevention 40, 776–786.

Donmez, B., Boyle, L., Lee, J.D., 2010. Differences in off-road glances: the effects on young drivers' performance. ASCE Journal of Transportation Engineering, Special Issue: Applications of Advanced Technologies in Transportation 136 (5), 403–409.

Donmez, B., Boyle, L.N., Lee, J.D., 2007. Safety implications of providing real-time feedback to distracted drivers. Accident Analysis and Prevention 39 (3), 581–590.

Donorfio, L.K.M., D'Ambrosio, L.A., Coughlin, J.F., Mohyde, M., 2009. To drive or not to drive, that isn't the question – the meaning of self-regulation among older drivers. Journal of Safety Research 40, 221–226.

Farah, H., Toledo, T., 2010. Passing behavior on two-lane highways. Transportation Research Part F: Traffic Psychology and Behaviour 13 (6), 355–364.

Folstein, M.F., Folstein, S.E., McHugh, P.R., 1975. Mini mental state: a practical method for grading the cognitive state for the clinician. Journal of Psychiatry Research 12 (3), 189–198.

Hakamies-Blomqvist, L., Mynttinen, S., Backman, M., Mikkonen, V., 1999. Age-related differences in driving: are older drivers more serial? International Journal of Behavioral Development 23 (3), 575–589.

Harbluk, J.L., Noy, Y.I., Trbovich, P.L., Eizenman, M., 2007. An on-road assessment of cognitive distraction: impacts on drivers' visual behavior and braking performance. Accident Analysis and Prevention 39 (2), 372–379.

Holland, C.A., Rabbitt, P.M., 1992. People's awareness of their age-related sensory and cognitive deficits and the implications for road safety. Applied Cognitive Psychology 6 (3), 217–231.

Holland, C.A., Rabbitt, P.M., 1994. The problems of being an older driver: comparing the perceptions of an expert group and older drivers. Applied Ergonomics 25 (1), 17–27.

Horberry, T., Anderson, J., Regan, M.A., Triggs, T.J., Brown, J., 2006. Driver distraction: the effects of concurrent in-vehicle tasks, road environment complexity and age on driving performance. Accident Analysis and Prevention 38 (1), 185–191.

Horrey, W.J., Lesch, M.F., Gabaret, A., 2008. Assessing the awareness of performance decrements in distracted drivers. Accident Analysis and Prevention 40 (2), 675–682.

Langford, J., Koppel, S., 2006. Epidemiology of older driver crashes – identifying older driver risk factors and exposure patterns. Transportation Research, Part F: Traffic Psychology and Behaviour 9 (5), 309–321.

Lavallière, M., Donmez, B., Reimer, B., Mehler, B., Klauber, K., Orszulak, J., et al., 2010. A field study assessing lane changing and lane choice across age and multiple levels of cognitive demand. In: Paper Presented at the Transportation Research Board 90th Annual Meeting, Washington, DC.

Matthews, G., Dorn, L., Hoyes, T.W., Davies, D.R., Glendon, A.I., Taylor, R.G., 1998. Driver stress and performance on a driving simulator. Human Factors 40 (1), 136–149.

Mehler, B., Reimer, B., Coughlin, J.F., 2012. Sensitivity of physiological measures for detecting systematic variations in cognitive demand from a working memory task: an on-road study across three age groups. Human Factors 54 (3), 396–412.

Mehler, B., Reimer, B., Coughlin, J.F., Dusek, J.A., 2009. The impact of incremental increases in cognitive workload on physiological arousal and performance in young adult drivers. Transportation Research Record: Journal of the Transportation Research Board 2138, 6–12.

Mehler, B., Reimer, B., Dusek, J., 2011. MIT AgeLab Delayed Digit Recall Task (n-back), MIT AgeLab White Paper Number 2011-3B. Massachusetts Institute of Technology, Cambridge, MA.

Michon, J.A., 1995. A critical view of driver behavior models: what do we know, what should we do? In: Evans, L., Schwing, R. (Eds.), Human Behavior and Traffic Safety. Plenum Press, New York, p. 520.

Olsen, E.C.B., Lee, S.E., Wierwille, W.W., Goodman, M.J., 2002. Analysis of distribution, frequency, and duration of naturalistic lane changes. In: Paper Presented at the Human Factors and Ergonomics Society 46th Annual Meeting, Baltimore, MD.

Pande, A., Abdel-Aty, M., 2006. Assessment of freeway traffic parameters leading to lane-change related collisions. Accident Analysis and Prevention 38 (5), 936–948.

Planek, T.W., Fowler, R.C., 1971. Traffic accident problems and exposure characteristics of the aging driver. Journal of Gerontology 26 (2), 224–230.

Reimer, B., 2009. Cognitive task complexity and the impact on drivers' visual tunneling. Transportation Research Record: Journal of the Transportation Research Board 2138, 13–19.

Reimer, B., Mehler, B., 2011. The impact of cognitive workload on physiological arousal in young adult drivers: a field study and simulation validation. Ergonomics 54 (10), 932–942.
Reimer, B., Mehler, B., Wang, Y., Coughlin, J.F., 2012. A field study on the impact of variations in short-term memory demands on drivers' visual attention and driving performance across three age groups. Human Factors 54 (3), 454–468.
Sodhi, M., Reimer, B., Llamazares, I., 2002. Glance analysis of driver eye movements to evaluate distraction. Behavior Research Methods Instruments and Computers 34 (4), 529–538.
Szlyk, J.P., Seiple, W., Viana, M., 1995. Relative effects of age and compromised vision on driving performance. Human Factors 37 (2), 430–436.
Wickens, C.D., 1984. Processing resources in attention. In: Davies, D.R., Parasuraman, R. (Eds.), Varieties of Attention. Academic Press, New York, NY.
Wickens, C.D., McCarley, J.S., 2008. Applied Attention Theory. CRC Press, New York.
Zeitlin, L.R., 1993. Subsidiary task measures of driver mental workload: a long-term field study. Transportation Research Record 1403, 23–27.
Zhou, H., Itoh, M., Inagaki, T., 2009. Effects of cognitive distraction on checking traffic conditions for changing lanes. In: Paper Presented at the Proceedings of the Human Factors and Ergonomics Society 53rd Annual Meeting, San Antonio, TX.

Multi-Modal Demands of a Smartphone Used to Place Calls and Enter Addresses during Highway Driving Relative to Two Embedded Systems

March 2015

Bryan Reimer
MIT AgeLab & New England University Transportation Center

Bruce Mehler
MIT AgeLab & New England University Transportation Center

Ian Reagan
Insurance Institute for Highway Safety

David Kidd
Insurance Institute for Highway Safety

Jonathan Dobres
MIT AgeLab & New England University Transportation Center

1005 N. Glebe Road, Suite 800
Arlington, VA 22201
+1 703 247 1500

iihs.org

Multi-Modal Demands of a Smartphone Used to Place Calls and Enter Addresses during Highway Driving Relative to Two Embedded Systems

Bryan Reimer*[1], Bruce Mehler[1], Ian Reagan[2], David Kidd[2], & Jonathan Dobres[1]

[1]*MIT AgeLab & New England University Transportation Center*
77 Massachusetts Avenue, E40-291, Cambridge, MA 02139, USA

[2]*Insurance Institute for Highway Safety*
1005 N. Glebe Road, Arlington, VA 22201, USA

*Corresponding author. Email: reimer@mit.edu
phone: (617) 452-2177 fax: (617) 258-7570

Funding: This work was supported by the Insurance Institute for Highway Safety and a grant from United States Department of Transportation's Region One University Transportation Center at MIT.

Disclosure Statement: The authors have no financial interest or benefit arising from the direct application of this research.

Multi-Modal Demands of a Smartphone Used to Place Calls and Enter Addresses during Highway Driving Relative to Two Embedded Systems

There is limited research on trade-offs in demand between manual and voice interfaces of embedded and portable technologies. Mehler et al. (2014a) identified differences in driving performance, visual engagement, and workload between two contrasting embedded vehicle system designs (Chevrolet MyLink and Volvo Sensus). The current study extends this work by comparing these embedded systems with a smartphone (Samsung Galaxy S4). None of the voice interfaces eliminated visual demand. Relative to placing calls manually, both embedded voice interfaces resulted in less eyes-off-road time than the smartphone. Errors were most frequent when calling contacts using the smartphone. The smartphone and MyLink allowed addresses to be entered using compound voice commands resulting in shorter eyes-off-road time compared with the menu-based Sensus but with many more errors. Driving performance and physiological measures indicated increased demand when performing secondary tasks relative to "just driving", but were not significantly different between the smartphone and embedded systems.

Practitioner summary: The findings show that embedded system and portable device voice interfaces place fewer visual demands on the driver than manual interfaces, but they also underscore how differences in system designs can significantly affect not only the demands placed on drivers but also the successful completion of tasks.

Keywords: Voice interface; visual demand; distraction; workload; human machine interface

1. Introduction

Since the dawn of the cellphone, there has been a debate concerning the dangers of phone use while driving. Studies have attempted to characterize the risks of phone use (Caird, Willness, Steel, & Scialfa, 2008; Collet, Guillot, & Petit, 2010; Dingus et al., 2006; Horrey & Wickens, 2006; McCartt, Hellinga, & Bratiman, 2006; McKnight & McKnight, 1993; Redelmeier & Tibshirani, 1997; Young & Schreiner, 2009), with studies using different methodologies and different measures producing widely varying estimates of risk and uncertainties about whether any elevated risk is explained by visual, manual, or cognitive attentional demands of cellphone use.

Several studies have examined safety-relevant events (e.g., near-crashes, traffic conflicts, crashes) using "naturalistic" driving data based on continuously monitoring drivers over weeks or even months. Recent studies (Fitch et al., 2013; Victor et al., 2014) have suggested that talking on a hand-held or hands-free phone may be risk-neutral or even protective. The reasons for this are not fully understood and appear counterintuitive considering consistent results from experimental research that indicate cellphone conversations delay drivers' reaction time and may affect other driving performance measures (Horrey & Wickens, 2006; Strayer & Drews, 2004; Strayer, Drews, & Crouch, 2006; Strayer, Drews, & Johnston, 2003). One well-considered issue that may reconcile this apparent conflict may be the phone's use by some drivers to combat monotony and fatigue under some circumstances (Atchley & Chan, 2011; Gershon, Shinar, Oron-Gilad, Parmet, & Ronen, 2011).

In contrast to studies of phone conversations using naturalistic driving data, studies using the same naturalistic driving data (Fitch, et al., 2013; Klauer et al., 2014; Victor, et al., 2014) have found that the visual-manual aspects of phone interaction such as dialing and texting are a significant source of increased risk of safety-relevant events. Further, studies using naturalistic driving data have repeatedly shown that various measures of drivers' eye deviations away from the roadway provide an indication of increased risk of safety relevant events (Klauer, Dingus, Neale, Sudweeks, & Ramsey, 2006; Victor, et al., 2014). It thus seems reasonable to hypothesize that systems placing fewer demands on a driver's visual attention to the roadway may be relatively safer than systems placing more demands on a driver's visual attention.

1.1 Research on voice interfaces

Voice-based interfaces are increasingly being integrated into vehicle infotainment systems and have been widely available in portable phones for a number of years. Voice-enabled interfaces have been proposed as a less demanding way to use phones, search for music, and enter navigational information (Chiang, Brooks, & Weir, 2005; Shutko, Mayer, Laansoo, & Tijerina, 2009). These systems have the potential to reduce, but not necessarily eliminate, the visual-manual demands associated with comparable visual-manual tasks (Chiang, et al., 2005; Mehler, Reimer, et al., 2014b; Owens, McLaughlin, & Sudweeks, 2011; Reimer, Mehler, Dobres, & Coughlin, 2013; Reimer et al., 2014; Shutko, et al., 2009).

Concerns have been raised about the cognitive demands of tasks that still remain with voice interfaces (Cooper, Ingebretsen, & Strayer, 2014; Reimer et al., 2013; Reimer, Mehler, Wang, & Coughlin, 2010, 2012; Strayer et al., 2013; Strayer, Turrill, Coleman, Ortiz, & Cooper, 2014). At the same time, several studies have found that self-reported workload, physiological arousal (e.g., heart rate), and other assessments of cognitive load (e.g., detection response task) are impacted to a lesser degree by voice interfaces than by visual-manual interfaces (Beckers et al., 2014; Chiang, et al., 2005; Mehler, Reimer, et al., 2014b; Munger et al., 2014; Owens, McLaughlin, & Sudweeks, 2010; Reimer, Mehler, Dobres, et al., 2013; Reimer, Mehler, et al., 2014; Shutko, et al., 2009). Not surprisingly, these studies also largely show that the demands of any secondary activity are greater than just driving alone.

1.2 Portable and embedded telematics use in the vehicle

Despite legislative efforts, phone usage in the vehicle remains high (Nurullah, Thomas, & Vakilian, 2013). Evidence on the effects of laws limiting drivers' phone use is mixed, so it is unclear whether the laws are achieving their intended purpose of reducing crashes (McCartt, Kidd, & Teoh, 2014). Given the prevalence of phone use in the vehicle, the uncertain effectiveness of laws curtailing their usage, and some research showing a divergence of risk associated with conversational aspects of phone use and dialing, it is imperative that we enhance our understanding of the trade-offs inherent in performing increasingly common in-vehicle tasks, using embedded vehicle or portable interfaces and across voice-based and visual-manual interfaces. While embedded systems increasingly allow drivers to complete phone and navigation tasks with manual and voice interfaces, many drivers prefer to use their smartphones for these tasks (Tison, Chaudhary, & Cosgrove, 2011). The reasons for this preference are not fully understood. However, familiarity with the smartphone, difficulties linking the smartphone to a vehicle through Bluetooth, the need to learn additional mental models for the vehicle embedded systems, and the desire for the latest technology are all likely contributing factors.

There is limited research on the trade-offs in demand between embedded vehicle systems and portable technologies. In the only field study that was identified, Owens et al. (2010) assessed driver behavior while using a production Ford SYNC voice interface for dialing and song selection compared with manual interaction through the drivers' own personal phone and portable music player. As the study was conducted several years ago, the assessment involved multiple antiquated technologies, such as 12-button numeric keypads and Apple iPods with a click-wheel. The study considered the demands of manually using the portable technologies for various tasks compared with the embedded voice system. It is unclear if the advantages observed for the embedded voice system over the manually used portable technologies (shorter task time; lower steering variance; lower maximum steering speed; shorter mean glance duration, lower total glance duration; fewer glances, lower maximum glance duration; and lower reported mental demand) for the tasks studied would generalize to a wider array of tasks such as phone contact calling and navigation entry and for more modern touchscreen smartphones.

Given the limited research comparing the demands of embedded vehicle telematics systems and smartphones, a field study was developed to assess driver behavior while engaging in contact calling and address entry tasks. Two vehicle embedded systems with divergent interface design approaches were selected for study based upon a hierarchical task analysis (Reagan & Kidd, 2013) of the steps required to use the visual-manual and voice-based interfaces to dial a contact stored in the embedded telematics systems. The selected vehicles were a 2013 Volvo XC60 with the Sensus infotainment system and a 2013 Chevrolet Equinox with MyLink. Considering the voice-based modes, the Sensus provided a menu-based voice interface that stepped through a series of menus and submenus. MyLink was designed around a “one-shot” voice interface where a single compound command could be used to execute most of a task. As a comparison to these embedded vehicle systems, a Samsung Galaxy S4 smartphone was mounted at a fixed location in each vehicle. The smartphone voice-based interface also supported a “one-shot” approach to entering commands and information about tasks analogous to that used by the MyLink voice interface. To fully categorize the benefits and drawbacks of the voice interfaces, contact calling tasks were also completed manually with the embedded systems and the smartphone.

1.3 Previous research and objective

A separate paper focuses on a comparison of the manual and voice interfaces of two embedded systems used to complete phone contact calling and voice navigation entry tasks (Mehler, Kidd, et al., 2014a). Overall, that report is consistent with previous literature (Chiang, et al., 2005; Mehler, Reimer, et al., 2014; Reimer, Mehler, Dobres, et al., 2013; Reimer, Mehler, et al., 2014; Shutko, et al., 2009) indicating that auditory-vocal interfaces can provide drivers with a means to decrease but not eliminate the time that their eyes are drawn away from the forward roadway when engaging in secondary tasks. In terms of the two embedded voice interfaces, the one-shot approach of MyLink showed distinct advantages in reduced task time and decreased visual demand compared with the menu-based Sensus system. The MyLink system was, however, limited by the accuracy of the voice recognition technology in the longer address entry tasks. In short, the Sensus menu-based voice interface led to longer interactions with more visual engagement, but maximized successful input of complex information compared with the MyLink’s one-shot approach.

The present work assessed the demands associated with the use of the manual and voice interfaces of the two markedly different in-vehicle embedded systems (Chevrolet Equinox with MyLink and Volvo XC60 equipped with Sensus) and a popular smartphone (Samsung Galaxy S4) mounted in the vehicle. While driving at highway speeds, participants used either the Chevrolet or Volvo embedded in-vehicle system and the mounted smartphone to perform phone contact calling and navigation system address entry tasks. Task demand was quantified across a range of variables including workload (heart rate, skin conductance, and self-report), visual engagement, and driving performance. Where applicable, significant differences between the two embedded vehicle systems and the smartphone are detailed. The embedded systems and the smartphone are compared for phone contact calling across both the manual and voice interfaces, while address entry was assessed only for the voice interface. The address entry task was not assessed using a manual interface as the perceived difficulty of manual address entry has led many manufacturers to block it while the vehicle is moving.

It was hypothesised that the newer cloud-based speech recognition technology in the smartphone would outperform the vehicles' embedded voice systems. Furthermore, given the design guidelines vehicle manufacturers use to limit attentional demand of in-vehicle systems (Driver Focus-Telematics Working Group, 2006; National Highway Traffic Safety Administration, 2013), the manual interfaces of the embedded vehicle systems were expected to be easier to use and less visually demanding for phone contact calling compared with the manual use of the smaller smartphone touchscreen.

2. Methods

2.1 Participants

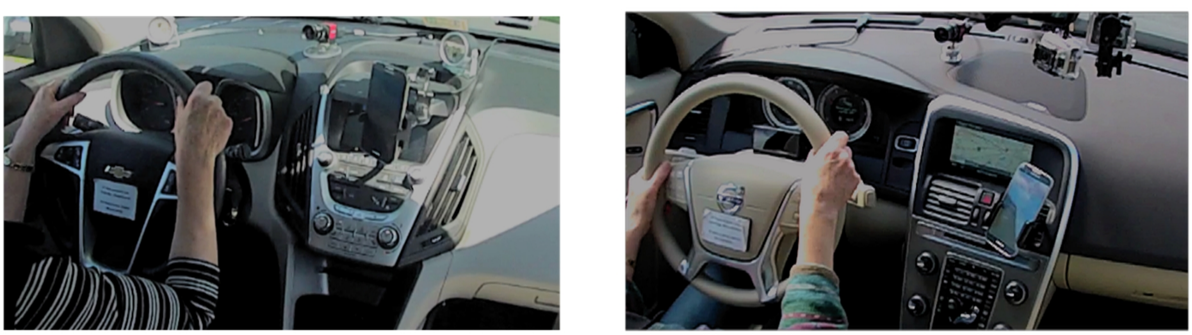
A sample of 122 relatively healthy and experienced drivers was recruited from the greater Boston area based upon responses to phone or on-line screening. Participants were required to be between the ages of 20 and 69, have been licensed for a minimum of 3 years, and self-report driving at least 3 times per week and being in relatively good health for their age. Also based on self-report, individuals were excluded if they had had a police-reported crash in the past year, had any of several specified medical conditions (e.g., a major illness resulting in hospitalization in the past 6 months, a diagnosis of Parkinson's disease, a history of stroke), or were taking medications (e.g., anti-convulsants, anti-psychotics, medications causing drowsiness) that might impair their ability to drive safely under the study conditions.

Forty-two participants were excluded from the analysis. Of these cases, six participated during protocol development; two were dropped due to protocol execution errors by a research associate; one was a participant who was unable to complete experimental tasks while driving (male 63 years of age); two indicated in the parking lot before the experiment began that they were unable or unwilling to complete experimental tasks (both female 64 years of age); four were cases where equipment failure occurred; five demonstrated unsafe driving behaviour; one did not meet the study criteria on closer examination; four were cases where the research associate noted unsafe or unusual weather or traffic conditions on the roadway; four had difficulty learning how to complete experimental tasks prior to driving (all males 45-64 years of age); one was a case where the smartphone did not consistently recognize the participant's voice, as determined during the experiment; one was a case where the MyLink system did not recognize the participant's voice in the parking lot prior to driving; one was a case where the MyLink system and smartphone did not recognize the participant's voice in the parking lot prior to driving; and one was excluded due to the research associate's discretion due to personal hygiene issues. A residual group of nine cases remained after it was confirmed that all the research matrix cells were filled with usable cases.

The final analysis sample of 80 cases was equally balanced across the two vehicles. The composition of the group in each vehicle was gender balanced and included an equal number of participants across the four age groups (18-24, 25-39, 40-54, 55 and older) specified in the National Highway Traffic Safety Administration's (2013) recommended guidance for assessing the extent of distraction from in-vehicle devices in the Visual-Manual Driver Distraction Guidelines for In-Vehicle Electronic Devices. Participant age did not vary significantly by gender or vehicle (*M* Female = 40.4 years, *M* Male = 40.3 years; *M* Chevrolet = 40.3 years, *M* Volvo = 40.4 years; both $F(1,79) = .949$) (see Mehler *et al.*, 2014a for detailed descriptive statistics). Recruitment procedures and the overall experimental protocol were approved by MIT's Committee on the Use of Humans as Experimental Subjects. Compensation of $75 was provided.

2.2 Apparatus

A 2013 Chevrolet Equinox equipped with the MyLink infotainment system and a 2013 Volvo XC60 equipped with the Sensus system were used. No modifications were made to the vehicle user interfaces. Smartphone connectivity was supported by pairing a Samsung Galaxy S4, model SCH-1545 (released March 2013) running Android 4.3 (Jelly Bean), to each vehicle's embedded system via the vehicle's Bluetooth wireless interface. A commercially available mount for the smartphone was attached to the center stack of each vehicle (Figure 1). As can be observed in the illustration, the distance and angle of reach to the smartphone varied between the two vehicles due to differences in the available mounting surfaces.



(a) (b)

Figure 1: Illustration of the smartphone mounting points in (a) the Chevrolet and (b) the Volvo

Both vehicles were instrumented with a customized data acquisition system for time synchronized recording of vehicle information from the controller area network (CAN) bus, a Garmin 18X Global Positioning system (GPS) unit, a MEDAC System/3™ physiological monitoring unit to provide EKG and skin conductance level (SCL) signals, video cameras, and a wide area microphone to capture driver speech and audio from the vehicle's speech system. The five video cameras provided views intended to capture the driver's face for primary glance behavior analysis, the driver's interactions with the vehicle's steering wheel and center console, the forward roadway (narrow and wide-angle images), and a rear roadway view. Data were captured at 10 Hz for the CAN bus and GPS, 30 Hz for the face and narrow forward roadway cameras, 15 Hz for the remaining cameras, and 250 Hz for the physiological signals to support EKG feature extraction for heart beat interval detection.

2.3 Secondary Tasks

2.3.1 Calling a phone contact

A phone list of 108 contacts was used for all phone calling tasks (see Mehler et al., 2014a for a more detailed description). Calling a phone contact was presented at two levels of difficulty. The easy tasks were calling a contact with only one phone number entry for that contact (Mary Sanders and Carol Harris). The hard tasks were calling a contact with two phone numbers (e.g., home and mobile). For these contacts (Pat Griffin on mobile and Frank Scott at work), the target phone was never the first listing so that simply requesting the contact name alone would not dial the correct number. The form of the easy task prompt was, ‘Your task is to call Mary Sanders. Begin.’ The form of the hard task prompt was, ‘Your task is to call Frank Scott at work. Begin.’ The contacts were the same across the manual and voice interface interactions so that any aspects/characteristics of a particular contact name that might influence the relative difficulty were constant (e.g., alphabetic location).

Calling a contact using the MyLink visual-manual interface began by locating and selecting the phone subsystem followed by selecting the alphanumeric bin (e.g., ABC, DEF) containing the target contact. The contact name was then selected from the list and a list of phone numbers were displayed, including a single number for the easy condition and multiple numbers for the hard condition. Calling a contact using the Sensus visual-manual interface required the user to select the phone subsystem and then scroll through the upper level of the contact list to the appropriate contact name using a rotary knob on the center console. The user then pressed an “OK” button to select the contact. When the contact had a single phone number (easy task), the call was initiated. For contacts with multiple numbers (hard task), a submenu listing the phone numbers for that contact was presented, and the rotary dial and “OK” button were used to locate and select the desired number. Manual calling a contact on the Samsung smartphone was initiated by turning the phone screen on by pressing the home button (a press button centered at the bottom of the phone). A “Contacts” icon appeared on the phone’s touch screen immediately above the home button; touching this button opened up the phone book and displayed a vertical listing of eight names in alphabetical order. Scrolling through the full list was carried out by sliding or swiping a finger up or down the screen surface. When the desired contact was visible, touching the entry brought up the contact page that displayed one or more phone numbers. A call was initiated by touching the desired number.

Calling a contact using the MyLink voice interface required very few steps. After pressing the push-to-talk button on the steering wheel, the driver could initiate both the easy and hard tasks in a single command string (e.g., “call Mary Sanders,” “call Pat Griffin on mobile”).

No confirmation step was required if the system had confidence in the identification of the selection. The Sensus voice interface closely mirrored the multi-level menu structure used in the manual interface. After pressing the push-to-talk button, the driver could issue the compound command 'Phone call contact' to access the phone list and then say the contact name (e.g., "Mary Sanders") following a prompt. A list of possible contacts would then appear on the display screen and the driver was asked to say a line number and then confirm the selection. In the case of the hard task where there were multiple phone numbers for the contact, a second level menu would appear showing the possible numbers. The driver selected from this listing verbally and confirmed the selection. The smartphone's S-Voice Drive feature (driving mode) was used for voice interaction. When this mode was enabled, tasks were initiated by pressing the home key twice and waiting through one of several variations of a standard greeting message ("Hello. I hope you're making the most of every day. When you need any help, say, 'Hi Galaxy.' "). The user then said "Hi Galaxy," waited for a tone indicating the system was ready to take a voice-command, and said "Call" followed by the desired contact name and number type if multiple entries were associated with the contact (e.g., "Pat Griffin on mobile").

Each phone number associated with a target contact connected with a voicemail recording that confirmed the contact identity and stated that the phone call could now be disconnected. If the target contact was not reached, the call connected to a voicemail indicating that the MIT AgeLab had been reached and the phone call could now be disconnected. This provided auditory confirmation to the participant and the research associate as to whether the target contact had been correctly selected or not.

2.3.2 Entering an address into a navigation system

During assessment, participants were asked to enter three addresses using the voice interface into each navigation system: 1) 177 Massachusetts Avenue, Cambridge, Massachusetts; 2) 293 Beacon Street, Boston, Massachusetts; and 3) their home address. The prompt was presented in the form, "Your task is to enter the destination address: 177 Massachusetts Avenue, Cambridge, Massachusetts. Begin." The first two addresses also were printed in large black text on a white card attached to the center of the steering wheel (see Figure 1) to minimize any cognitive load of needing to memorize and hold the address in memory during the duration of the interaction with the navigation system. The card was in place throughout the drive so that participants were exposed to the addresses for a minimum of 40 minutes prior to being asked to enter them into the system.

Voice address entry with MyLink was initiated by pressing the "push-to-talk" button and saying the command "navigation." After prompting the driver for a navigation command, the system accepted various commands to begin destination entry including "destination address," "enter address," and simply "address." The complete address was then entered as a single verbal string (e.g., "177 Massachusetts Avenue, Cambridge, Massachusetts"). If the system was confident in identification, there was no confirmation step, and navigation instructions were initiated unless multiple potential targets were identified; in this case, a list of addresses were presented auditorially to the user to select from. With Sensus, the command "navigate go to address" was used to select address entry. Then Sensus prompted the user for each part of the address in individual steps (i.e., city name, street name, and street number). The user was prompted to confirm or correct their entry by voice after each step by verifying the visual information displayed on the navigation interface in the centre stack. Once the address was entered correctly, the driver was prompted to say "finish" and then say "enter destination" to initiate navigation. If the system identified multiple potential targets, a list of options was shown on the center stack display screen and the system prompted the driver to "say a line number or say not on list." The smartphone used the Google Maps application for address entry. The task was initiated by pressing the home key twice, waiting through the greeting message, saying "Hi Galaxy," and waiting for a tone indicating the system was ready to take a voice-command. The driver then said "Navigate to" followed by the address (e.g., "177 Massachusetts Avenue, Cambridge, Massachusetts') in a single verbal string. A tone sounded and the system said "I will navigate you to" followed by its interpretation of the address string. The Google Maps application then displayed a map on screen, and audio instructions for navigation became active. Participants were instructed during training to cancel the application by touching the back button repeatedly until the home screen reappeared.

2.4 Experimental design

Participants were randomly assigned to one of the two vehicles. As represented schematically in Figure 2 and further detailed in section 2.5 on procedure, participants were presented with the phone contact calling tasks to be undertaken using voice-based and manual interfaces and with the address entry task using the voice-based navigation interfaces. For each participant, tasks were performed using one of the embedded vehicle systems and a smartphone. Within each vehicle group, random assignment was made to either an "embedded vehicle system" or a "smartphone" first condition. Within each condition, random assignment determined whether voice-based or manual phone contact calling was presented first. Consequently, any advantage of being presented with the same contact to dial a multiple times was balanced across the interfaces. The address entry tasks were always presented between the two forms of phone calling for a particular system.

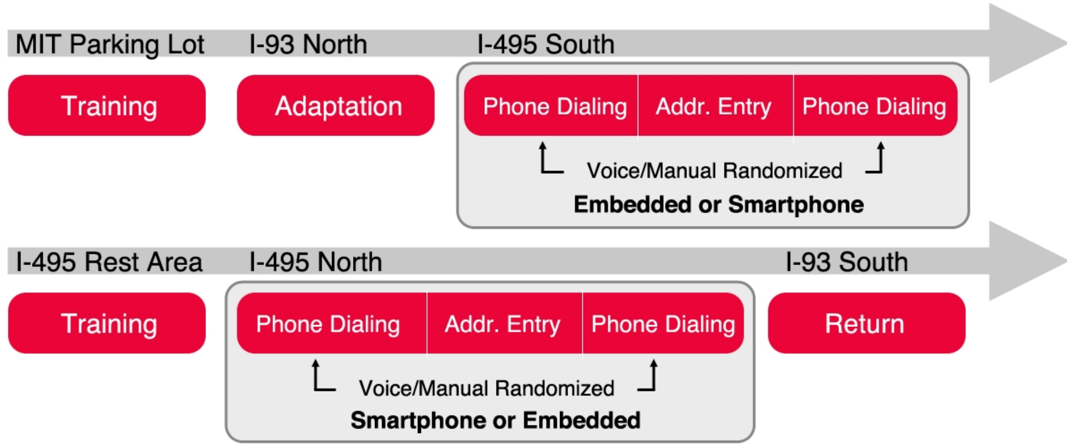


Figure 2: Schematic representation of the experimental design. Half of the participants interacted with the embedded vehicle systems on I-495 South and half with the smartphone. Device type (embedded or smartphone) was reversed for the I-495 North segment so that all participants experienced both types.

In summary, across six distinct task periods, each participant was presented a total of 22 secondary tasks, 11 during the southbound segment using either the embedded or smartphone system (four manual phone contact calling trials, three address entry trials, and four voice calling trials) and then the same 11 tasks during the northbound segment using the alternate device.

2.5 Procedure

Participants reviewed and signed an informed consent, and a structured interview was conducted to confirm eligibility. Information on participants' demographic characteristics, attitudes toward driving, and experience with technology was gathered by questionnaire; an explanation of the workload rating scale was provided; and physiological sensors were attached. An EKG recording was obtained using a modified lead II configuration that placed the negative lead just under the right clavicle, the ground just under the left clavicle, and the positive lead over the lowest left rib. Gold-plated skin conductance sensors were attached with medical grade paper tape on the underside of the outer segment of each of the two middle fingers of the left hand.

After being escorted to the research vehicle, participants were instructed on how to adjust the seat and mirrors, and, where necessary, how to operate the keyless ignition system. Participants were trained in the parking lot in the use of the embedded or smartphone system to which they were assigned for the first half of the drive. Training began with manual phone contact calling, followed by voice phone contact calling, and then by voice destination address entry. For the embedded vehicle systems, following the approach taken in Reimer et al. (2013) and Mehler et al. (2014b), the default factory-setting configurations for the vehicle voice interfaces were used, and participants were given guidance on the use of short-cut command options to reduce the number of steps required to complete tasks. As an example of a shortcut, to use the voice interface in the Sensus system, calls could be placed by first saying the command

'Phone," waiting for a response, and saying 'Call Contact." During training, participants where told "Calls can be placed by speaking the command 'Phone Call Contact;' you can also use the shorter command, 'Call Contact.' " The remainder of the training interaction then focused on the shorter version. Participants with the Volvo Sensus system were taken through the voice calibration procedure, which is intended to tune the voice recognition system to the participants' pronunciation based on a set of command relevant words; the Chevrolet MyLink system did not have this feature. For the portion of the study using the Samsung Galaxy smartphone, the smartphone was placed in the dashboard mount. Orientation and training for both embedded systems and the smartphone consisted of recorded instructions to provide consistency, supplemented with guidance by a research associate to clarify details and answer questions. Participants were encouraged to repeat tasks until they felt comfortable to proceed. The orientation/training period typically ranged between 15 and 30 minutes, with a mean of approximately 20 minutes.

Participants then drove the vehicle on actual roadways in and around the greater Boston area. A driving adaptation period of approximately 30 minutes took place prior to the start of the experiment and consisted of approximately 10 minutes of urban driving from MIT to interstate highway I-93 and approximately 20 minutes north on I-93 to I-495. For the portions used in this study, I-495 is a divided interstate that is largely surrounded by forest with three traffic lanes in each direction with lane widths of 15 feet (3.62 m). The posted speed limit is 65 mph (104.6 kph).

Presentation of the secondary tasks with the first assigned system interface (smartphone or embedded system) occurred while driving south on I-495 (see Figure 2). At the end of this southbound segment, a break was taken at a highway rest stop where participants completed workload and other ratings for the tasks just completed. They were then trained on the alternate interface (smartphone or embedded) on the same set of secondary tasks. Assessment of the alternate interface then took place during the second half of the drive as participants proceeded north on I-495, and participants completed the workload and other ratings for the second set of tasks on their return to the MIT parking lot.

Smartphone assessments were always conducted with the device secured in the dashboard mount. The phone was always removed from the mount for the segment of driving involving the assessment of the embedded vehicle systems. Most participants took approximately 35 to 40 minutes to drive each segment (north and south) (70 to 80 minutes combined).

The difficulty of the phone calling tasks was presented within each voice or manual period in the following order: easy, easy, hard, hard. This was intended to provide participants additional familiarity with the interface before assessing the harder task trials. Between individual trials, there was an interval of 30 seconds after the research associate recorded the completion of a task and the recorded instructions began for the next. A separation period of at least 3 minutes was provided following the end of one group of related tasks and the next period (e.g., between phone calling and address entry). During address entry trials, the navigation application was left active after an address entry for approximately 30 seconds prior to the driver being prompted by recorded instructions to cancel the application. This allowed for clear separation between behaviors associated with entering an address and canceling the application. The total contact time for the study including intake and debrief was typically about 4 hours. Participants were instructed several times (in the written consent form, by recorded instructions, and through direct prompting by the research associate in the vehicle) that at all times during the driving portion of the study, priority should be given to safe driving.

2.6 Dependent measures

Mehler et al. (2014a) provides background and detail on the outcome measures collected. In brief, subjective workload was assessed using a single global rating per secondary task type on a 0 (low) to 10 (high) scale that allowed for half-interval ratings (21 points). The instruction set and scale have been demonstrated to produce ratings consistent with relative rankings of global scores obtained using the NASA Task Load Index (Beckers, et al., 2014; Hart, 2006; Munger, et al., 2014). Physiological measures (heart rate and skin conductance level) were recorded as they have been shown to be sensitive to changes in objectively graded levels of working memory load (Mehler, Reimer, & Coughlin, 2012; Mehler, Reimer, Coughlin, & Dusek, 2009; Reimer & Mehler, 2011) and other demands during driving (Brookhuis & de Waard, 2001; Collet, Salvia, & Petit-Boulanger, 2014; Yang, Reimer, Mehler, & Dobres, 2013). Task time and major wheel reversals (gap size > 3 degrees) were computed based upon CAN recordings and time stamps provided by the data acquisition system. Vehicle speed and the standard deviation of speed were calculated based on GPS values and expressed as percentage change from baseline driving. Visual demand metrics (mean duration of individual (single) glances, the percentage of glances per participant greater than 2.0s, and the total time a participant glanced away from the forward road scene) were computed based upon manually reduced eye data (see description below). Finally, task error rates originating from the user and system are reported.

2.7 Data analysis

2.7.1 Subjective workload, behavioral and physiological measures

Baseline driving reference periods consisted of 2 minutes of just driving prior to a recorded audio message indicating that a new task period was about to start (see Figure 1). There were six such baseline periods per participant on the I-495 portion of the drive, and a seventh 2-minute reference was recorded on I-93 south on the return to MIT (14 minutes total). Values for relevant metrics were calculated, and the mean values across the baseline periods were used as a baseline, “just driving” reference.

Task completion time was calculated as the time between the end of a task prompt and successful completion or failure of the task. Instantaneous heart rate was computed by locating R-wave peaks in the EKG signal and determining the inter-beat intervals using software developed at the MIT AgeLab. In line with existing standards (Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology, 1996), automated detection results were visually reviewed and misidentified and irregular intervals manually corrected. Skin conductance was post processed using another MIT-developed package designed to remove high-frequency noise in the signal, following procedures detailed in Reimer and Mehler (2011), and allowing for manual editing of motion artifacts.

Eye glance measures were quantified following ISO standards (ISO 15007-1, 2002; ISO 15007-2, 2001) with a glance to a region of interest defined to include the transition time to that object. In the case of manual coding of video images, the timing of glance is labeled from the first video frame illustrating movement to a "new" location of interest to the last video frame prior to movement to a "new" location. Glance data for this study were manually coded using software, now available as open source (Reimer, Gruevski, & Coughlin, 2014), that allowed for rapid frame-by-fame review and coding. Each task period of interest was independently coded by two evaluators. Discrepancies between the evaluators (the identification of conflicting glance targets, missed glances, or glance timings that differed by more than 200ms) were mediated by a third researcher. The taxonomy and procedures for this coding methodology were initially proposed in Smith, Chang, Glassco, Foley, & Cohen (2005) and detailed further in Reimer, Mehler, Dobres, et al. (2013, Appendix G).

Statistical analyses were performed in R (R Core Team, 2014). Owing to the non-normal distribution of the data and/or the use of ratio data (percentages) for several dependent measures, in many cases non-parametric statistics such as the Wilcoxon signed rank test and the Friedman test were used (similar to the t-test and repeated-measures ANOVA, respectively). For multifactorial analyses, repeated-measures ANOVA by ranks are presented. These tests have been shown to be more robust against Type I error in cases where data are non-normal (Conover & Iman, 1981; Friedman, 1937).

For analysis of the contact phone calling tasks, the primary statistical tests assumed a model in which the vehicle driven (Chevrolet or Volvo) was a between-subjects factor, and device (embedded or smartphone) and modality (manual or voice) were within-subjects factors, resulting in a 2 × (2 × 2) mixed design. Analysis of the destination address entry task assumed a model in which the vehicle driven was a between-subjects factor and device (embedded or smartphone) a within-subject factor, resulting in a 2 × 2 mixed design. Since the focus of the analysis was to examine the effects of different device types (and input modality in the case of phone calling), the vehicle driven was included to control for the effects of vehicle in the model, but main effects and interaction of the vehicle factor are reported only where the effect of vehicle results in notable differences between the primary variables of interests. In these cases, for comparative purposes, an alternate version of the results is presented controlling for vehicle (i.e., considering the impact on a variable relative to an average of the two vehicles utilized in the study). As noted earlier, comparisons of the embedded vehicle systems are fully detailed in Mehler et al. (2014a).

3. Results

Findings are presented first for the phone contact calling tasks and then for the destination address entry tasks. In considering the phone tasks, 'modality' refers to the overt method of interface interaction (manual or voice) and device refers to the embedded vehicle systems versus the smartphone. As noted earlier, in selected cases, references to differences observed between the two vehicles and their specific embedded system are provided to enhance the understanding of effects related to smartphone use.

3.1 Phone Contact Calling

3.1.1 Self-reported workload

Workload ratings differed significantly by device ($F(1, 77) = 9.68$, $p = .003$) and input modality ($F(1, 77) = 113.57$, $p < .001$). In addition, there was a significant interaction between these factors ($F(1, 77) = 11.20$, $p = .001$). As illustrated in Figure 3, workload ratings for the voice tasks were lower than for the manual calling tasks; however, the reduction in workload associated with voice calling relative to manual calling was significantly greater for the embedded systems than the smartphone.

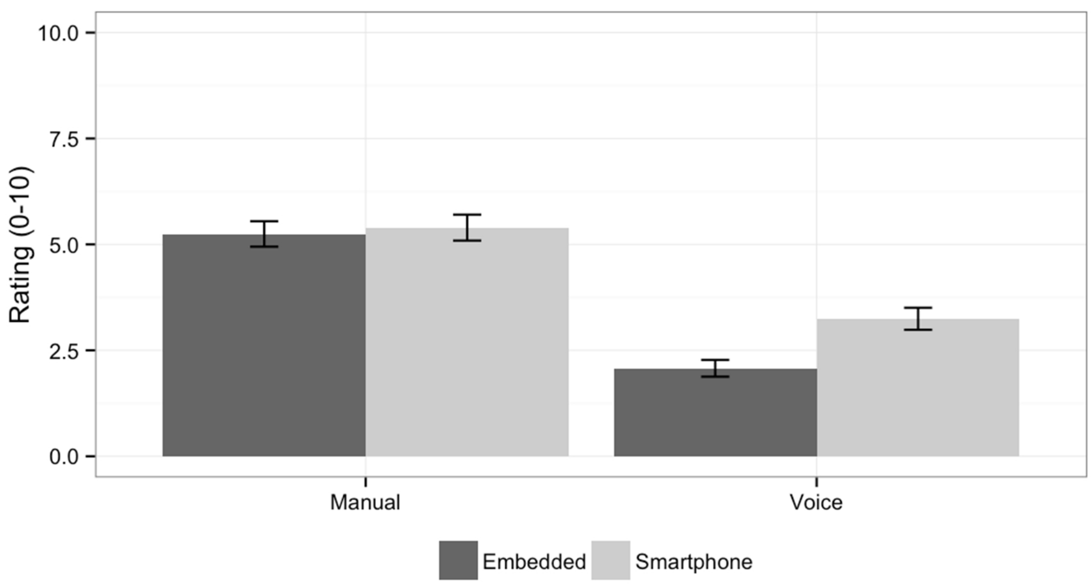


Figure 3: Mean self-reported workload ratings for phone calling tasks on a scale of 0 (low) to 10 (high) by device and interface type. Error bars represent ±1 standard error.

3.1.2 *Task Completion Time*

Phone task completion time was affected by a significant interaction between vehicle driven and device type ($F(1, 78) = 42.69$, $p < .001$), as well as a significant three-way interaction between vehicle driven, device type, and input modality ($F(1, 78) = 13.66$, $p < .001$). Therefore, the three-way interaction was decomposed by vehicle driven to gain a clearer understanding of these factors' effects on phone task completion time.

3.1.2.1 Chevrolet

In the Chevrolet, phone task completion time varied significantly by device ($F(1, 39) = 149.66$, $p < .001$) and modality ($F(1, 39) = 53.62$, $p < .001$). The significant interaction between device and modality ($F(1, 39) = 120.98$, $p < .001$) reveals that phone task completion times for manual interactions were similar for the embedded device and smartphone, but varied considerably when the voice interface was used (Figure 4). Smartphone voice calling tasks took more than twice as long to complete compared with the Chevrolet's embedded vehicle interface.

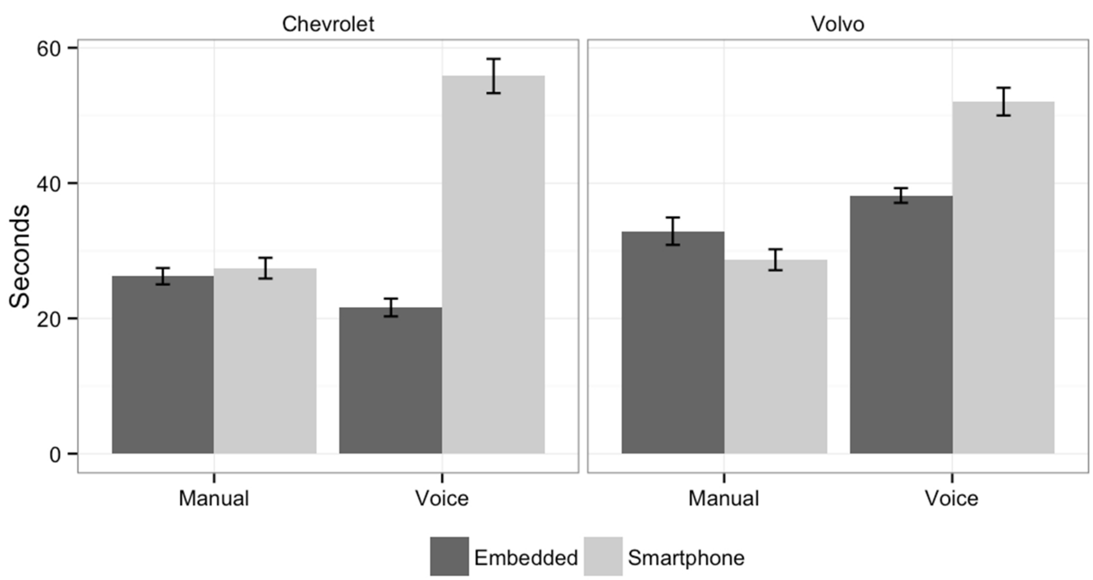


Figure 4: Mean task completion time in seconds for phone calling by vehicle, device, and interface type. Error bars represent ±1 standard error.

3.1.2.2 Volvo

In looking at the phone task completion times in the Volvo, the main effects were consistent with those for task completion times in the Chevrolet. Tasks completed with the smartphone interface took longer to complete compared with the embedded system ($F(1, 39) = 13.01$, $p < .001$). Voice contact calling tasks required significantly more time to complete compared with their manual equivalents ($F(1, 39) = 100.44$, $p < .001$). However, in contrast with the Chevrolet, the statistical interaction between device and modality in the Volvo ($F(1, 39) = 44.70$, $p < .001$) points to a more complex relationship between the visual-manual and voice interfaces across devices. Consistent with the Chevrolet, when using voice interfaces, tasks took longer to complete with the smartphone compared with the embedded system. When using the manual interfaces, however, the opposite pattern was observed.

Thus, the three-way interaction reflects varying differences in task completion time using each system's voice interface relative to the manual interface and different relationships among each embedded manual interface relative to the smartphone manual interface. On average, voice contact calling tasks took less time to complete relative to manual contact calling tasks using the Chevrolet embedded system, but took longer using the Volvo embedded system and even longer using the smartphone voice interface. In addition, there was a small but significant increase in the time participants in the Volvo took to complete the manual contact calling tasks with the embedded interface compared with the smartphone interface, while there was no comparable difference for manual contact calling in the Chevrolet.

3.1.3 *Physiological Metrics*

Heart rate increased during phone task periods relative to baseline single task driving ($V = 710$, $p < .001$, Wilcoxon test of mean task heart rate vs. baseline heart rate), rising by a mean of 1.9%. The average percentage change in heart rate was not significantly different between devices (*M* embedded = 2.18% [*SE* = 0.37%], *M* smartphone = 1.64% [*SE* = 0.40%]; $F(1, 78) = 0.83$, $p = .366$) or between input modalities (*M* manual = 1.90% [*SE* = 0.41%], *M* voice = 1.92% [*SE* = 0.35%]; $F(1, 78) = 0.000$, $p = .953$), nor did these factors interact significantly ($F(1, 78) = 2.09$, $p = .152$).

As was the case with heart rate, mean skin conductance levels increased significantly during phone contact calling relative to baseline driving ($V = 38$, $p < .001$). Skin conductance changes were significantly affected by modality ($F(1, 72) = 4.50$, $p = .037$), with skin conductance levels rising over baseline driving by 13.4% (*SE* = 1.69%) during manual tasks versus 9.6% (*SE* = 1.54%) during voice tasks. Skin conductance changes were not affected by device (M embedded = 11.34% [SE = 1.56%], M smartphone = 11.71% [SE = 1.68%]; $F(1, 72) = 0.02$, $p = .900$), and no interaction between device and modality was observed ($F(1, 72) = 0.11$, $p = .745$).

3.1.4 *Glance Behavior Metrics*

The effect of vehicle driven on mean off-road single glance duration was significant ($F(1, 78) = 4.06$, $p = .047$). Mean off-road glance duration was also significantly affected by the interaction between vehicle driven and modality ($F(1, 78) = 11.09$, $p < .001$), vehicle driven and device ($F(1, 78) = 10.62$, $p = .002$), and a three-way interaction between vehicle driven, device, and modality ($F(1, 78) = 18.40$, $p < .001$). The pattern of mean single off-road glance durations for participants who drove the Chevrolet indicate that both the manual and voice interfaces of the embedded system had shorter mean single off-road glances than the manual and voice interfaces of the smartphone, respectively (Figure 5). Participants in the Chevrolet also had shorter mean single off-road glances when using the embedded and smartphone voice interfaces compared with the manual interfaces.

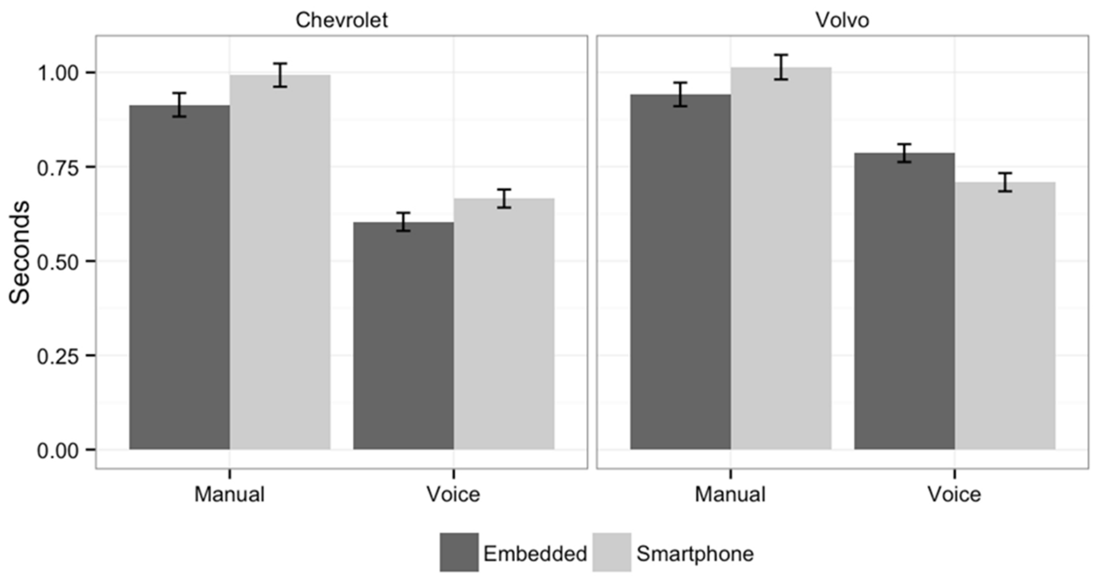


Figure 5: Mean single off-road glance duration during phone contact calling by vehicle, device, and interface type. Error bars represent ±1 standard error.

Similar to participants in the Chevrolet, participants who used Volvo's Sensus had reductions in mean single off-road glances when using the embedded and smartphone voice interfaces relative to their manual counterparts (Figure 7). However, whereas the MyLink voice interface had a greater reduction in mean single off-road glances relative to the smartphone interface, use of the Sensus voice interface was associated with greater mean single off-road glance durations than the smartphone voice interface. In addition to the three-way interaction, there was a significant main effect of device (*M* embedded = 0.81s [*SE* = 0.02s]; *M* smartphone = 0.85s [*SE* = 0.02s]; $F(1, 78) = 8.41$, $p = .005$), a significant main effect of modality (*M* manual = 0.97s [*SE* = 0.02s]; *M* voice = 0.69s [*SE* = 0.01s]; $F(1, 78) = 426.64$, $p < .001$), and a significant interaction between the two factors ($F(1, 78) = 18.40$, $p < .001$). Mean single off-road glance duration was significantly shorter when using the embedded device compared with the smartphone during manual calling tasks (*M* embedded = 0.93s, *M* smartphone = 1.00s) but was similar when using the voice interfaces of the devices (*M* embedded = 0.69s, *M* smartphone = 0.69s).

Long duration glances were significantly affected by input modality (*M* manual = 3.5% [*SE* = 0.4%]; *M* voice = 0.5% [*SE* = 0.1%]; $F(1, 78) = 52.98$, $p < .001$), but not by device ($F(1, 78) = 2.67$, $p = .106$). Furthermore, a significant device by modality interaction was observed ($F(1, 78) = 4.41$, $p = .039$). Specifically, the use of the voice interfaces resulted in a similar low percentage of long duration glances for both embedded (*M* = 0.5%, *SE* = 0.3%) and smartphone interfaces (*M* = 0.4%, *SE* = 0.3%), whereas for manual interfaces, the smartphone showed a higher frequency of long duration glances than the embedded interfaces (4.0% [*SE* = 0.60%] and 2.9% [*SE* = 0.52%], respectively).

There was a significant interaction between vehicle driven, device, and modality for total off-road glance time: ($F(1, 78) = 5.84$, $p = .018$). The details of this effect are discussed below. In more general terms, total off-road glance time was significantly affected by device ($F(1, 78) = 40.59$, $p < .001$), with the embedded systems requiring less glance time compared with the smartphone (Figure 6). There was also a significant effect of modality ($F(1, 78) = 81.27$, $p < .001$), with voice interfaces requiring less glance time compared with manual interfaces. These factors also interacted significantly ($F(1, 78) = 56.28$, $p < .001$) so that the use of the embedded voice interfaces required the least total off-road glance time. Based on these analyses, it is clear that within each vehicle, the embedded voice interface was associated with less off-road glance time compared with the embedded manual interface, and with both the voice and manual interfaces of the smartphone. However, the relative reduction in off-road glance time when using the voice interface was different for the two embedded systems, and this resulted in the three-way interaction which is discussed below.

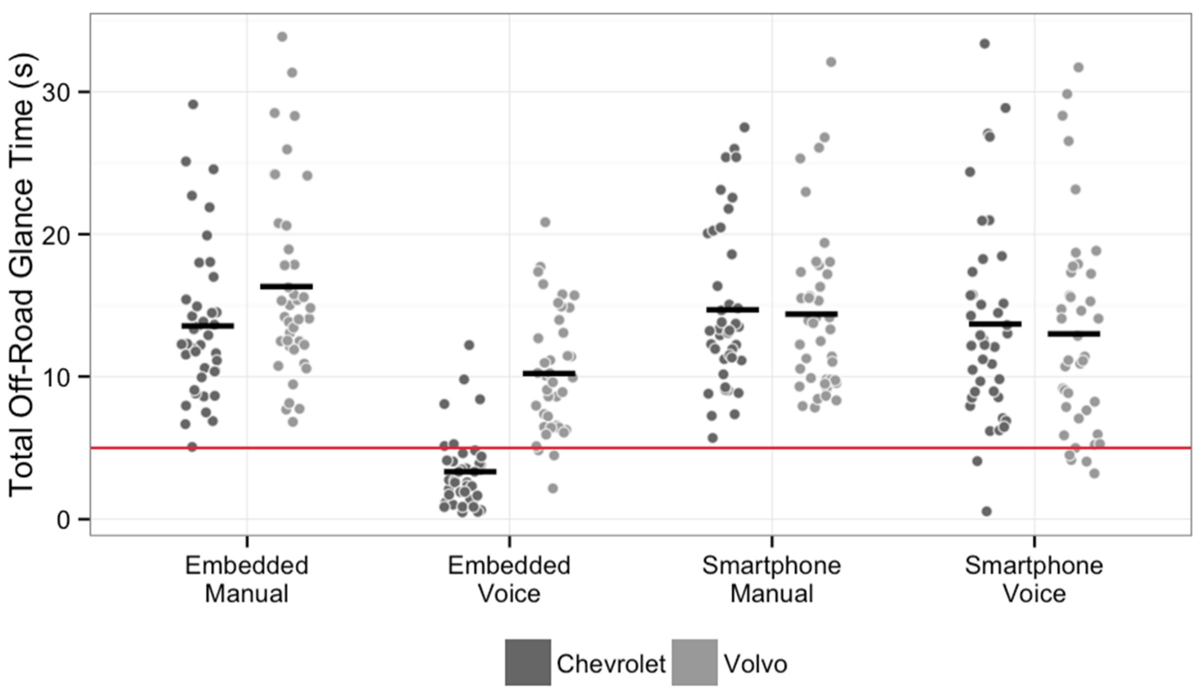


Figure 6: Cumulative off-road glance times for each phone dialing task by vehicle driven. Points indicate total off-road glance time for each participant and have been jittered horizontally to minimize overlap. The short line segments indicate mean total off-road glance time for each group. The long horizontal line represents 5 seconds of total off-road glance time.

The magnitude of the difference in total off-road glance time for the phone calling tasks between the embedded system and smartphone interfaces varied across the two vehicles studied. While these data could be examined by vehicle driven at the group level, a consideration of the data at the individual participant level provides a more comprehensive view of differences between the systems. As illustrated in Figure 6, almost all participants required a minimum of

5 seconds (indicated by the solid horizontal line) of cumulative off-road glance time to complete the manual phone calling tasks and voice-based smartphone calling tasks.

For the calling tasks, individual participants' total off-road glance durations during use of the Volvo embedded voice interface cluster below all of the manual interfaces and the smartphone voice-based interface. In addition, most drivers' off-road glance durations during the use of the Volvo voice interface were more than 5 seconds. In contrast, most of drivers' off-road glance durations were less than 5 seconds when using the Chevrolet's embedded voice interface for contact calling, with just seven participants requiring more than 5 seconds of total off-road glance time. Thus, while the embedded voice interfaces of both vehicles showed advantages in total off-road glance time, the effect was most pronounced in the Chevrolet implementation (and hence the three-way interaction stated above).

3.1.5 *Vehicle Control Metrics*

Participants decreased their driving speed by a mean of 2.4% during phone calling task periods relative to baseline driving only ($V = 2832$, $p < .001$). A main effect of modality appeared ($F(1, 78) = 6.84$, $p = .011$); manual calling tasks ($M = -3.1\%$, $SE = 0.54\%$) were associated with a greater decrease in speed compared with voice calling tasks (M = -1.6%, $SE = 0.36\%$). Device used (embedded or smartphone) did not affect speed ($F(1, 78) = 0.66$, $p = .418$), nor was there an interaction with modality ($F(1, 78) = 2.97$, $p = .089$).

Standard deviation of vehicle speed decreased significantly during phone calling task periods compared with baseline driving ($V = 3232$, $p < .001$); the percentage point difference between means was 36.6%. The percentage change in the standard deviation of vehicle speed during task periods relative to baseline driving was significantly affected by device (*M* embedded = -41.1% [*SE* = 2.27%], *M* smartphone = -32.1% [*SE* = 2.20%]; $F(1, 78) = 9.58$, $p = .003$), but not by input modality (*M* manual = -38.2% [*SE* = 2.29%], *M* voice = -35.0% [*SE* = 2.22%]; $F(1, 78) = 1.29$, $p = .260$). These factors interacted significantly ($F(1, 78) = 28.61$, $p < .001$); the percentage reduction in standard deviation of speed was greater during voice contact calling relative to manual contact calling when using the embedded devices (*M* voice = -45.3%, *SE* = 2.69%; *M* manual = -36.9%, *SE* = 3.60%) but not when using the smartphone (*M* voice = -24.7%, *SE* = 3.16%; *M* manual = -39.5%, *SE* = 2.84%). However, this pattern of results only reflects task performance using the Chevrolet embedded system, which was significantly different from the Volvo system, as indicated by a three-way interaction between vehicle driven, device type, and input modality ($F(1, 78) = 8.88$, $p = .004$). There was no difference in the percentage change in standard deviation of vehicle speed when using the Volvo embedded system for manual or voice contact calling, but the percentage reduction in this measure was greater when manual contact calling with the smartphone compared with voice (Figure 7).

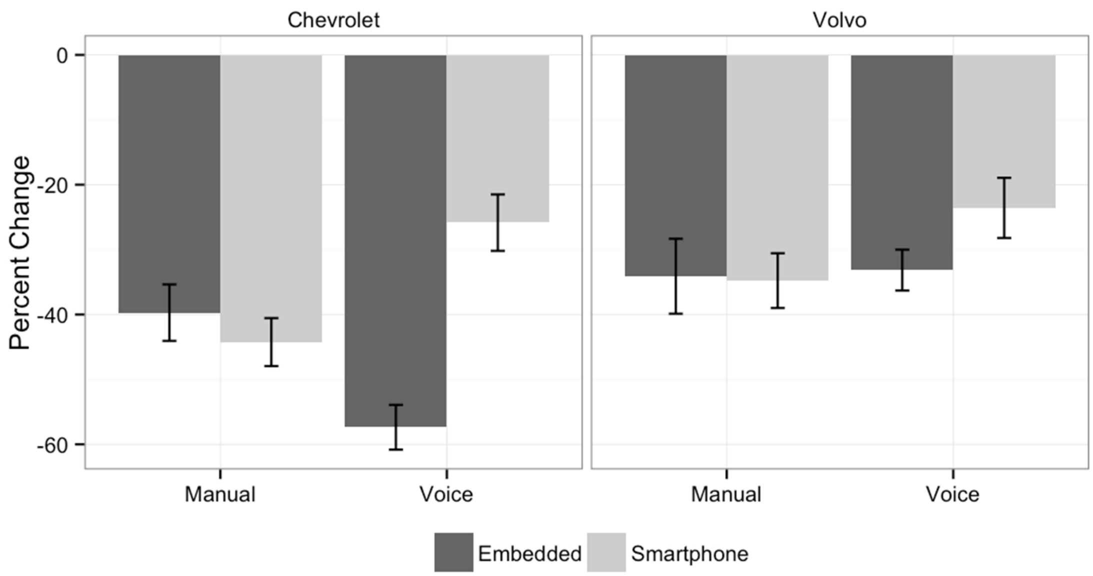


Figure 7: Mean percentage change from baseline of standard deviation of vehicle speed during phone contact calling by vehicle, device, and interface type. Error bars represent ±1 standard error.

Standard deviation of speed and mean speed were not significantly correlated (tests of mean speed per participant vs. mean standard deviation per participant (R = 0.005, p = .96). This indicates that the two metrics are independent of one another or, in other words, that standard deviation of speed does not decrease simply as a function of decreasing mean speed.

Major steering wheel reversal rates increased by 33.4% during phone task periods compared with baseline driving (V = 422, p < .001). Major steering wheel reversal rates were not affected by device type (F(1, 78) = 0.14, p = .714), but the rate of major steering wheel reversals was significantly higher during manual phone contact calling tasks than voice (M manual = 15.7/min [SE = 1.00/min], M voice = 14.2/min [SE = 0.96/min]; F(1, 78) = 5.41, p = .023). The interaction between device type and modality was not significant (F(1, 78) = 1.03, p = .313).

3.2 ***Destination address entry***

Voice entry tasks were performed with only the voice interfaces of the embedded devices and the smartphone. Table 1 summarizes the results of ANOVA-by-ranks tests for a main effect of device during the address entry tasks while controlling for the effect of the vehicle driven for the various dependent measures.

Table 1: Mean values (standard errors in parentheses) and statistical results of tests for an effect of device (embedded or smartphone) during address entry tasks with voice interfaces.

Measure	F	p	Embedded *M* (*SE*)	Smartphone *M* (*SE*)
Self-Reported Workload	1.4	0.248	3.07 (0.3)	3.46 (0.3)
Task Completion Time	177.7	0.001	73.64 (1.8)	48.32 (1.7)
Percent Change in Heart Rate	2.9	0.091	1.45 (0.5)	2.85 (0.6)
Percent Change in SCL	1.4	0.249	7.33 (2.4)	11.86 (2.2)
Mean Single Glance Duration	36.4	0.001	0.78 (0.0)	0.71 (0.0)
Percentage of Off-Road Glances > 2.0sec	11.3	0.001	1.15 (0.2)	0.55 (0.1)
Total Off-Road Glance Time	29.8	0.001	18.42 (1.0)	13.49 (0.8)
Percent Change in Mean Velocity	2.3	0.137	-0.22 (0.4)	0.27 (0.6)
Percent Change in SD of Velocity	6.4	0.013	-19.94 (3.7)	-27.03 (4.1)
Major Wheel Reversals	0	0.858	12.33 (1.2)	12.94 (1.4)

The smartphone voice interface resulted in significantly less visual demand during destination address entry tasks than the embedded system voice interfaces, as indicated by significantly lower mean off-road single glance duration, percentage of long duration off-road glances, and total off-road glance time for the smartphone compared with the average of the two embedded vehicle systems Table 1. In addition, task completion time was significantly shorter for the smartphone voice interface than the embedded voice interfaces. The percentage reduction in standard deviation of speed during task periods relative to baseline driving was significantly greater during use of the smartphone voice interface compared with the embedded voice interfaces. No other comparisons reached statistical significance.

There were significant interactions of vehicle driven and device type on measures of total off-road glance time ($F(1, 78) = 25.31$, $p < .001$), total task time ($F(1, 78) = 25.70$, $p < .001$), and change in variability of speed from baseline driving ($F(1, 78) = 8.41$, $p = .005$). No other significant interactions were observed. Total off-road glance time was significantly longer when using the Volvo's embedded voice interface to perform the navigation tasks (*M* = 22.5s, *SE* = 1.43s) compared with the other voice interfaces (*M* Volvo smartphone = 12.9s [*SE* = 1.07s], *M* Chevrolet embedded = 14.3s [*SE* = 1.22s], *M* Chevrolet smartphone = 14.1s [*SE* = 1.29s]). Task completion times were similar when using the smartphone voice interface regardless of vehicle driven (*M* Chevrolet = 51.3s, *SE* = 2.65s; *M* Volvo = 45.3s, *SE* = 2.14s) and longer than when using both embedded system voice interfaces; however, task completion times using the Volvo embedded voice interface (*M*=80.6s, *SE* = 1.71s) were much longer than when using the Chevrolet's (*M* = 66.7s, *SE* = 2.85s). The percentage change in standard deviation of vehicle speed relative to just driving was similar when performing the navigation task using the embedded voice system or smartphone voice system in the Chevrolet (*M* embedded = -29.5% [SE = 4.58%], *M* smartphone = -28.7% [*SE* = 4.03%]); however, a greater percentage reduction in standard deviation of speed was observed when using the smartphone voice system to perform the navigation task in the Volvo compared with Volvo's embedded voice system (*M* embedded = -10.4% [*SE* = 5.46%], *M* smartphone = -25.4% [*SE* = 7.27%]).

3.3 *Error analysis & interaction characterization*

Errors made during completion of the phone contact calling and address entry tasks were analyzed in two ways. First, task trials were classified as error-free or, for trials where an error occurred, as a trial with a user error or a trial with a system error. User errors were instances where a participant spoke an incorrect voice command that resulted in the task not moving forward or progressing incorrectly, selected incorrect manual input, or when the research assistant provided assistance. System errors were instances where a participant issued a correct voice command that was understood by the research associate in the vehicle but was misinterpreted by the voice recognition system. If both a system error and user error occurred in the same trial, then the trial was categorized as a user error regardless of the total number of user or system errors that occurred. Thus, system errors are likely underrepresented in this analysis method.

Each trial also was categorized based on the degree of difficulty a participant encountered when completing the task. Individual trials were categorized as 1) error-free, 2) completed with backtracking, 3) completed with one instance of assistance from the research associate, 4) completed with more than one instance of assistance from the research associate, or 5) as a failure. Backtracking was defined as instances where the system did not recognize or misinterpreted a command and provided another opportunity for the voice command to be entered; this included instances where the participant restarted the task without aid from the research associate. Backtracking could also occur because a participant recognized that they made an error (such as giving a wrong street name) and used an option provided by the system to correct the error. The research associate in the vehicle provided assistance to the participant when he judged that a participant was not going to progress through a task on his/her own. One or more instances of researcher assistance were provided to participants to increase the chance that the task in a given trial was completed successfully. This support was provided to mitigate the participant's frustration and to allow for monitoring whether correction of simple misunderstandings or forgetting of commands resolved initial problems in using the systems while driving. A trial was categorized as a failure if the participant had to restart the task more than twice, failed to progress in the task despite receiving assistance from the research associate, or cases where the system or user executed the task incorrectly. Both methods of error coding were completed by two members of the research staff who independently evaluated each trial. One staff member was the research associate in the vehicle during the drive and the other was an associate who reviewed video and audio recordings of the drive. A third staff member mediated discrepancies.

3.3.1 *Errors: Contact calling*

The contact calling trials performed with the manual interfaces of the embedded systems or smartphone were more often error-free (91%) than the contact calling trials performed using the voice interfaces (84%). With voice calling tasks, there were markedly more trials performed using the smartphone that ended in failure, with backtracking, or that required assistance from the research associate compared with trials performed using the embedded vehicle systems (Table 3). An error was coded for about 25% of the trials completed using the smartphone voice interface, whereas an error was coded for 7.5% of the trials completed using the embedded system voice interfaces. The percentage of manual calling trials that were error-free when participants used an embedded system was slightly higher (93%) than when the manual interface of the smartphone was used (88%). Of the manual calling trials completed with the smartphone where an error occurred, the majority of errors were backtracking (Table 2).

Table 2. Percentage of contact calling trials in each error category for each interface modality and device.

Modality	Device	Error-free	Backtracking	One instance of assistance	More than one instance of assistance	Failure
Manual interface	Embedded systems	92.8	2.8	2.2	0.9	1.3
	Smartphone	88.1	8.4	2.8	0.3	0.3
Voice interface	Embedded systems	92.5	1.6	4.1	1.3	0.6
	Smartphone	75.9	6.6	6.6	5.0	6.0

Note. Percentages are based on 320 total trials for each row except for the smartphone voice interface which had 1 trial that could not be categorized (n=319).

Table 3 provides the number of phone calling trials with a user error and the number with a system error across device and modality. Overall, there were nearly twice as many trials with a user error (8.4%) as trials with a system error (4.4%). About 87% of trials with a system error (48 out 55) occurred when using the smartphone's voice interface.

Table 3. Percentage of contact calling trials without errors, system errors, or user errors for each interface modality and device.

Modality	Device	Error-free	System error	User error
Manual interface	Embedded systems	92.8	0.0	7.2
	Smartphone	88.1	0.0	11.9
Voice interface	Embedded systems	92.5	2.2	5.3
	Smartphone	75.9	15.0	9.4

Note. Percentages are based on 320 total trials for each row except for the smartphone voice interface which had 1 trial that could not be categorized (n=319).

3.3.2 *Errors: Navigation entry*

The percentage of address entry trials that were coded as error-free was smaller for trials performed with the smartphone or embedded system in the Chevrolet (59%) than with the smartphone or embedded system in the Volvo (86%) (Table 4). When considering the smartphone and embedded systems separately, the percentage of trials that were error-free was substantially lower among Chevrolet drivers who entered addresses using MyLink (49.2%) than among Chevrolet drivers using the smartphone (69.2%). In contrast, a somewhat larger proportion of address entry trials were error-free when Volvo drivers used Sensus (89.2%) compared with the smartphone (82.5%). The percentage of address entry trials that ended with failures was highest when drivers used MyLink (20%) and lowest with Sensus (1.7%), with 14.2% of trials using the smartphone ending in failure in the Chevrolet and 9.2% of these trials ending in failure in the Volvo.

Table 4. Percentage of address entry trials coded in each error category for each interface modality and device.

Vehicle	Device	Error-free	Backtracking	One instance of assistance	More than one instance of assistance	Failure
Chevrolet	Embedded system	49.2	7.5	10.8	12.5	20.0
	Smartphone	69.2	6.7	6.7	3.3	14.2
Volvo	Embedded system	89.2	3.3	2.5	3.3	1.7
	Smartphone	82.5	5.0	3.3	0.0	9.2

Note. Percentages are based on 120 total trials for each row.

In general, the percentage of navigation tasks with a system error (19.0%) was greater than the percentage of trials with user errors (8.5%). A system error was noted in almost three times as many address entry trials among participants driving the Chevrolet (28%) compared with trials completed by participants driving the Volvo (9.6%). Trials with a user error were most commonly recorded among participants who were using the Chevrolet MyLink to enter addresses compared with when they used the smartphone and participants using the smartphone or Sensus in the Volvo (Table 5).

Table 5. Percentage of address entry trials coded without errors or with a system or user error for each vehicle and device.

Vehicle	Device	Error-free	System error	User error
Chevrolet	Embedded system	49.2	31.7	19.2
	Smartphone	69.2	25.0	5.8
Volvo	Embedded system	89.2	4.2	6.7
	Smartphone	82.5	15.0	2.5

Note. Percentages are based on 120 total trials for each row.

4. Discussion

4.1 Phone contact calling

Consistent with patterns observed in previous research on infotainment systems (Chiang, et al., 2005; Mehler, Reimer, et al., 2014b; Owens, et al., 2011; Reimer, Mehler, Dobres, et al., 2013; Reimer, Mehler, et al., 2014; Shutko, et al., 2009) , the voice-based methods of phone contact calling were associated with lower self-reported workload ratings and lower visual demand (mean single glance duration, percentage of glances longer than 2 seconds, and total eyes-off-road time) compared with the manual methods. The present analysis extends other work by showing that this pattern of results holds for the Samsung smartphone as well as for the embedded vehicle systems studied. Further, while heart rate as an arousal measure did not show an advantage for either modality, skin conductance levels were consistent with lower workload, on average, during voice-based calling compared with manual calling.

For phone contact calling, the apparent advantages for the voice interfaces were greater with the embedded vehicle systems than with the smartphone across a number of metrics. On average, self-reported workload and total eyes-off-road time were lower using the embedded systems' voice interfaces than the smartphone voice interface and their manual counterparts. Thus, pairing a smartphone with a vehicle's embedded system and using the embedded system's voice interface may reduce workload and visual demand. In this regard, it is worth noting that the smartphone was mounted during these evaluations. Considering the demands and risks associated with picking up and handling a phone (Famer, Klauer, McClafferty, & Guo, 2014; Fitch, et al., 2013; Klauer, et al., 2014) one might anticipate additional benefits for embedded systems relative to smartphones that are not mounted. It should be noted that a single smartphone was examined, and the findings may not apply to other smartphones with different design approaches and different voice recognition technology. Mehler et al.'s (2014a) study illustrated how different embedded vehicle system designs have varying effects on driver workload and visual scanning, and presumably these measures also would vary across different smartphone interface designs. The smartphone in this study was selected because the Android platform had the largest market presence and the screen size was larger, rather than for its specific interface design characteristics.

Returning again to the broader question of using a smartphone or embedded vehicle system for contact calling, the total number of errors was higher when using the smartphone. This held for both the manual and voice methods of calling. One of the factors for the higher system error rate for the voice interface in the smartphone may be related to the positioning of the phone. When mounted on the dashboard, the microphone was farther away from drivers than if they were holding it in their hand, possibly affecting sound quality and voice recognition. To the extent that this is the case, it would suggest that the characteristics of the microphones used in the cars were more effective for this application. Similarly, the touch screen interface on the phone was optimized for handheld operation. Reaching for and touching smaller icons on the smartphone might explain some of the higher user errors in the manual smartphone mode versus the manual mode for the embedded systems. Additional characteristics relative to the voice interface on the smartphone are considered below in the context of the destination address entry task.

Considering the primary driving performance metrics, there were no significant differences by device type (embedded or smartphone) in terms of the degree of speed reduction or in the extent to which major steering wheel reversal rates increased during task periods. Thus, no relative advantage for embedded systems or the smartphone was apparent in these measures. Steering wheel reversal rates were higher during manual calling than voice calling, which is consistent with increased competition for manual resources between the driving task and the secondary task during manual calling relative to voice calling. Voice calling using the embedded systems was associated with a larger reduction in the standard deviation of speed relative to baseline driving than manual calling. The general reduction in speed variability during task periods relative to baseline driving may reflect drivers shifting their attention away from vehicle control to interacting with the embedded system. However, it is not clear why greater reductions in speed variability were observed with voice calling than manual calling given that voice calling presumably interfered with the driving task less than manual calling; this should be a topic for future research.

The time taken to make phone calls with the smartphone voice interface was significantly longer than for the embedded voice interfaces. This is likely related, at least in part, to the initial greeting message played each time when the driving mode of the voice interface of the Samsung Galaxy was engaged, followed by the need to say "Hi Galaxy" prior to being able to issue a voice-command. This added time and a layer to each task that was not present in either embedded system.

Experience with the Samsung's driving mode also highlighted the potentially dynamic nature of smartphone-based user interfaces. Software updates were blocked on the study phones to ensure a consistent user experience across participants. Nonetheless, one phone was inadvertently allowed to update and the voice-interface was modified as a result. For purposes of the study, the update was rolled back. However, exploring the updated software revealed significant changes from the software version tested during the study. For example, the extended greeting message that had previously played each time driving mode was activated at the beginning of a task was no longer present, removing this time-consuming aspect of the earlier interface. It is an open question as to how a driver learns about and adapts to such system upgrades.

4.2 Destination Address Entry

While a number of advantages were observed for the embedded systems for voice-based phone contact calling compared with the smartphone, a somewhat different picture appears in comparing the smartphone and embedded system interfaces for voice-based destination address entry. While self-reported workload and increases in heart rate and skin conductance were all nominally higher for the smartphone interfaces than for the embedded systems, these differences were not statistically significant. In contrast, mean single glance duration, the percentage of long duration glances, and total eyes-off-road time were all significantly greater for the embedded systems. Broadly considered, it appears that the Samsung smartphone voice-based system for the destination address entry task provided a less visually demanding engagement than the embedded alternatives.

The apparent advantage in visual demand for the smartphone voice interface for destination address entry must be tempered somewhat in evaluating net advantage for the two system types when errors are taken into account. In this regard, differences in system implementation features and possible differences in the vehicle environment may interact to impact the overall task experience. As detailed in Mehler et al. (2014a), the segmented approach to address entry used by the Volvo Sensus system (breaking voice input into independent chunks for city, street name, and street number) took more time and involved greater total eyes-off-road time compared with the Chevrolet MyLink system using a one-shot approach, but Sensus was associated with fewer system recognition errors. A similar difference appears when comparing the embedded Sensus implementation with the Samsung smartphone implementation, which also provided a one-shot address entry; specifically, the one-shot approach reduced visual demand when successful, but it also had a higher error rate. Comparing the one-shot address entry of the Samsung and MyLink voice interfaces, the smartphone had fewer errors.

Interestingly, system-based error rates for voice-based address entry in the smartphone were higher for participants who drove in the Chevrolet Equinox than for those who drove in the Volvo XC60. While this could be a chance finding, the research staff believed that ambient road noise was higher in the Equinox, and that this might have impacted voice recognition. In consideration of this hypothesis, an assessment of the background sound levels in each of the vehicles at 65 mph highway speed was conducted. Three sound readings were recorded in each vehicle at the respective mounting position of the smartphones. The average of the three readings indicated that the ambient noise levels in the Chevrolet Equinox were louder than those in the Volvo XC60 at the 125 Hz band (65dBA Equinox; 62 dBA XC60) and the 2000 Hz band (62.6dBA Equinox; 60dBA XC60), suggesting that ambient sound level differences between the vehicles could have contributed to the observed differences in voice recognition errors for both the embedded systems and the smartphone.

Ambient noise, and the earlier reported differences in speed reported by each vehicle's CAN bus and different mounting positions of the smartphones (driven by the physical layout of each vehicle's dashboard), illustrate some of the complexities of conducting inter-vehicle comparisons. While field experiments allow for observing driver use of technologies in a real-world environment provide valuable data, they have limitations, and the findings of field experiments are best understood when considered together with other methods to develop a comprehensive understanding of the technologies.

4.3 Limitations

The study sample was comprised of novice users. Some of the drawbacks noted (e.g., higher error rates) for the Smartphone and embedded system voice interfaces may not be observed among actual owners who have more familiarity using the voice commands or menu structure. Additionally, the visual demands observed with novice user interactions with the voice interfaces may not generalize to experienced users who know the sequence of commands or pace of turn-taking when completing tasks with the voice interface.

Another limitation is that participants may have felt compelled to perform the contact calling and address entry tasks in situations where they normally would not. The task instructions and research assistants repeatedly emphasized that participants should not perform a task if they felt unsafe or would not engage in the task during personal driving; however, no participants who went on-road declined to engage in a task.

Additionally, the extent to which effects associated with the dependent measures analyzed translate into safety risk are unclear. As emphasized in Reimer and Mehler (2013), such driver and visual performance data are informative concerning the attentional demand characteristics of the interface tasks, rather than necessarily being predictive of risk to drivers who are operating their own vehicles. Cognitive workload is inherently difficult to evaluate and was assessed indirectly as a component of self-reported workload and through peripheral physiological indices of arousal (heart rate and skin conductance).

The use of the same phone contacts and destination addresses across the different interfaces could be questioned. However, the use of the same entry tasks for each interface removed the necessity to characterize and identify different addresses that had equal levels of difficulty with regard to speech complexity. Counterbalancing the order of interface assessment across the sample should have controlled for any issue of presentation order. Moreover, the nature of phone calling and address entry tasks is such that it is likely that many drivers will call the same contacts and enter the same destinations into a navigation system relatively frequently.

Finally, it is unknown how manual entry might differ for embedded systems or the smartphone in performance of navigation tasks. However, it should be noted that the vehicles tested in this study locked out manual address entry when the vehicle was moving, and given concerns with the safety of entering an address while driving, ethical considerations would have prevented assessing manual performance on this task in the current field setting. Similarly, the set of secondary tasks assessed in this study were limited to placing phone calls and entering addresses. Whether similar patterns would be observed for tasks more complex than contact calling, such as sending voice-based text messages, is an area for future research.

5. Conclusions

Overall, the results suggest that there are benefits and drawbacks to voice interface technology in the smartphone relative to two embedded voice systems. While the smartphone largely outperformed the embedded vehicles' voice systems in task time and various metrics of visual engagement for the destination address entry task, it showed a smaller reduction in total off-road glance time when placing calls using voice input compared with manual input. In terms of manual interactions, results are also mixed. Average task completion time for contact calling using the smartphone was shorter than when using Sensus but no different from MyLink. The evidence is converging, however, that voice interfaces offer a less visually demanding way to access and input information than manual alternatives. In so much as drivers choose to engage in contact calling, the embedded vehicle voice interfaces would appear to be the most advantageous method of the ones considered in the current study. In contrast, the relative benefits of the voice interfaces of the embedded vehicle or the smartphone voice interfaces for destination address entry are not as clear.

The complex relationships between outcome measures need to be weighed when developing systems and considering their potential impact on safety. Clearly, a system that is capable of performing an operation with minimal demand on the driver is desirable. However, a brittle system that has difficulty performing requested operations without errors may be no more advantageous than a system that places more demands on attentional resources yet performs flawlessly.

6. Acknowledgments

Primary support for this work was provided by the Insurance Institute for Highway Safety (IIHS), with additional support provided through a grant by the US DOT's Region I New England University Transportation Center at MIT. Acknowledgment is gratefully extended to Peter Hamscher, Alex Hruska, Martin Lavallière, Alea Mehler, Hale McAnulty, Mauricio Muñoz, Lauren Parikhal, Anthony Pettinato, Adrian Rumpold, and Andrew Sipperley for their contributions in protocol development, data collection, reduction, and manual scoring. Appreciation is also extended to Adrian Lund, Anne McCartt, and David Zuby for review and comment on the manuscript.

7. References

Atchley, P., & Chan, M. (2011). Potential benefits and costs of concurrent task engagement to maintain vigilance: A driving simulator investigation. *Human Factors, 53*(3), 3-12.

Beckers, N., Schreiner, S., Bertrand, P., Reimer, B., Mehler, B., Munger, D., & Dobres, J. (2014). Comparing the demands of destination entry using Google Glass and the Samsung Galaxy S4. *Proceedings of the Human Factors and Ergonomics Society 58th Annual Meeting*, Chicago, IL.

Brookhuis, K. A., & de Waard, D. (2001). Assessment of drivers' workload: Performance and subjective and physiological indexes. In P. A. Hancock & P. A. Desmond (Eds.), *Stress, Workload, and Fatigue* (pp. 321-333). Mahwah, NJ: Lawrence Erlbaum Associates.

Caird, J. K., Willness, C. R., Steel, P., & Scialfa, C. (2008). A meta-analysis of the effects of cell phones on driver performance. *Accident Analysis & Prevention, 40*(4), 1282-1293.

Chiang, D. P., Brooks, A. M., & Weir, D. H. (2005). Comparison of visual-manual and voice interaction with contemporary navigation system HMIs. *SAE Technical Paper 2005-01-0433*.

Collet, C., Guillot, A., & Petit, C. (2010). Phoning while driving II: a review of driving conditions. *Ergonomics, 53*(3), 602-616.

Collet, C., Salvia, E., & Petit-Boulanger, C. (2014). Measuring workload with electrodermal activity during common braking actions. *Ergonomics, 57*(6), 886-896.

Conover, W. J., & Iman, R. L. (1981). Rank transformations as a bridge between parametric and nonparametric statistics. *American Statistician, 35*(3), 124-129.

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.

Dingus, T. A., Klauer, S. G., Neale, V. L., Petersen, A., Lee, S. E., Sudweeks, J.D., Perez, M.A., Hankey, J., Ramsey, D.J., Gupta, S., Bucher, C., Doerzaph, Z.R., Jermeland, J., & Knipling, R.R. (2006). The 100-Car Naturalistic Driving Study, Phase II – Results of the 100-Car Field Experiment: U.S. Department of Transportation, National Highway Traffic Safety Administration.

Driver Focus-Telematics Working Group. (2006). Statement of Principles, Criteria and Verification Procedures on Driver Interactions with Advanced In-Vehicle Information and Communication Systems, Version 2.0: Alliance of Automotive Manufacturers.

Famer, C. M., Klauer, S. G., McClafferty, J. A., & Guo, F. (2014). Relationship of near-crash/crash risk to time spent on a cell phone while driving. Arlington, VA: Insurance Institute for Highway Safety.

Fitch, G. A. S., S.A., Guo, F., McClafferty, J., Fang, Y., Olson, R. L., Perez, M. A., et al. (2013). The impact of hand-held and hands-free cell phone use on driving performance and safety-critical event risk (Report No. DOT HS 811 757). Washington, DC: National Highway Traffic Safety Administration (NHTSA).

Friedman, M. (1937). The use of ranks to avoid the assumption of normality implicit in the analysis of variance. *Journal of the American Statistical Association, 32*(200), 675-701.

Gershon, P., Shinar, D., Oron-Gilad, T., Parmet, Y., & Ronen, A. (2011). Usage and perceived effectiveness of fatigue countermeasures for professional and nonprofessional drivers. *Accident Analysis & Prevention, 43*(3), 797-803.

Hart, S. G. (2006). NASA-task load index (NASA-TLX); 20 years later. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 904-908.

Horrey, W. J., & Wickens, C. D. (2006). Examining the impact of cell phone conversations on driving using meta-analytic techniques. *Human Factors, 48*(1), 196-205.

ISO 15007-1. (2002). Road vehicles - Measurement of driver visual behaviour with respect to transport information and control systems - Part 1: Definitions and parameters. Geneva, Switzerland: International Standards Organization.

ISO 15007-2. (2001). Road vehicles - Measurement of driver visual behaviour with respect to transport information and control systems - Part 2: Equipment and procedures. Geneva, Switzerland: International Standards Organization.

Klauer, S. G., Dingus, T. A., Neale, V. L., Sudweeks, J. D., & Ramsey, D. J. (2006). The impact of driver inattention on near-crash/crash risk: an analysis using the 100-Car Naturalistic Driving Study Data (Report No. DOT HS 810 594). Washington, DC: United States Department of Transportation, National Highway Traffic Safety Administration.

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.

McCartt, A. T., Hellinga, L. A., & Bratiman, K. A. (2006). Cell phones and driving: review of research. *Traffic Injury Prevention, 7*(2), 89-106.

McCartt, A. T., Kidd, D. G., & Teoh, E. R. (2014). Driver cellphone and texting bans in the United States: Evidence of Effectiveness. *Annals of Advances in Automotive Medicine, 58*, 99-114.

McKnight, A. J., & McKnight, A. S. (1993). The effect of cellular phone use upon driver attention. *Accident Analysis and Prevention, 25*(3), 259-265.

Mehler, B., Kidd, D., Reimer, B., Reagan, I., Dobres, J., & McCartt, A. (2014a). Multi-modal assessment of on-road demand of voice and manual phone calling and voice navigation entry across two embedded vehicle systems. Arlington, VA: Insurance Institute for Highway Safety.

Mehler, B., Reimer, B., & Coughlin, J. F. (2012). Sensitivity of physiological measures for detecting systematic variations in cognitive demand from a working memory task: an on-road study across three age groups. *Human Factors, 54*(3), 396-412.

Mehler, B., Reimer, B., Coughlin, J. F., & Dusek, J. A. (2009). The impact of incremental increases in cognitive workload on physiological arousal and performance in young adult drivers. *Proceedings of the Transportation Research Board of The National Academies*, Washington, DC.

Mehler, B., Reimer, B., Dobres, J., McAnulty, H., Mehler, A., Munger, D., & Coughlin, J.F. (2014b). 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.

Munger, D., Mehler, B., Reimer, B., Dobres, J., Pettinato, A., Pugh, B., et al. (2014). A simulation study examining smartphone destination entry while driving. *Proceedings of the 6th International Conference on Automotive User Interfaces and Interactive Vehicle Applications*, Seattle, WA.

National Highway Traffic Safety Administration. (2013). Visual-Manual NHTSA Driver Distraction Guidelines for In-Vehicle Electronic Devices (Docket No. NHTSA-2010-0053). Washington, DC: U.S. Department of Transportation National Highway Traffic Safety Administration (NHTSA).

Nurullah, A. S., Thomas, J., & Vakilian, F. (2013). The prevalence of cell phone use while driving in a Canadian province. *Transportation Research Part F: Traffic Psychology and Behaviour, 19*, 52-62.

Owens, J. M., McLaughlin, S. B., & Sudweeks, J. (2010). On-road comparison of driving performance measures when using handheld and voice-control interfaces for mobile phones and portable music players. *SAE International Journal of Passenger Cars – Mechanical Systems, 3*(1), 734-743.

Owens, J. M., McLaughlin, S. B., & Sudweeks, J. (2011). Driver performance while text messaging using handheld and in-vehicle systems. *Accident Analysis & Prevention, 43*(3), 939-947.

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

Reagan, I. J., & Kidd, D. G. (2013). Using heirarchical task analysis to compare four vehicle manufacturers' infotainment systems. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, San Diego, CA, 1495-2599.

Redelmeier, D. A., & Tibshirani, R. J. (1997). Association between cellular-telephone calls and motor vehicle collisions. *New England Journal of Medicine, 336*(7), 453-458.

Reimer, B., Gruevski, P., & Coughlin, J. F. (2014). MIT AgeLab Video Annotator. Cambridge, MA. Retrieved from TBD posting of open sourced code is in progress.

Reimer, B., & Mehler, B. (2011). The impact of cognitive workload on physiological arousal in young adult drivers: a field study and simulation validation. *Ergonomics, 54*(10), 932-942.

Reimer, B., Mehler, B., Dobres, J., & Coughlin, J. F. (2013). The effects of a production level “voice-command” interface on driver behavior: reported workload, physiology, visual attention, and driving performance (Technical Report 2013-17a). Cambridge, MA: MIT AgeLab.

Reimer, B., Mehler, B., Dobres, J., McAnulty, H., Mehler, A., Munger, D., & Coughlin, J.F. (2014). Effects of an ‘expert mode’ voice command system on task performance, glance behavior & driver physiology. *Proceedings of the 6th International Conference on Automotive User Interfaces and Interactive Vehicle Applications*, Seattle, WA.

Reimer, B., Mehler, B., McAnulty, H., Munger, D., Mehler, A., Perez, E. A. G., Manhardt, T., & Coughlin, J.F. (2013). A preliminary assessment of perceived and objectively scaled workload of a voice-based driver interface. *Proceedings of the 7th International Driving Symposium on Human Factors in Driver Assessment, Training, and Vehicle Design*, Bolton Landing, NY, 537-543.

Reimer, B., Mehler, B., Wang, Y., & Coughlin, J. F. (2010). The impact of systematic variation of cognitive demand on drivers’ visual attention across multiple age groups. *Proceedings of the 54th Annual Meeting of the Human Factors and Ergonomics Society*, San Francisco, CA, 2052-2056.

Reimer, B., Mehler, B., Wang, Y., & Coughlin, J. F. (2012). A field study on the impact of variations in short-term memory demands on drivers' visual attention and driving performance across three age groups. *Human Factors, 54*(3), 454-468.

Shutko, J., Mayer, K., Laansoo, E., & Tijerina, L. (2009). Driver workload effects of cell phone, music player, and text messaging tasks with the Ford SYNC voice interface versus handheld visual-manual interfaces *SAE Technical Paper 2009-01-0786*. doi: 10.4271/2009-01-0786

Smith, D. L., Chang, J., Glassco, R., Foley, J., & Cohen, D. (2005). Methodology for capturing driver eye glance behavior during in-vehicle secondary tasks. *Transportation Research Record: Journal of the Transportation Research Board, 1937*(1), 61-65.

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.

Strayer, D. L., & Drews, F. A. (2004). Profiles in driver distraction: Effect of cell phone conversation on younger and older drivers. *Human Factors, 46*(4), 640-649.

Strayer, D. L., Drews, F. A., & Crouch, D. J. (2006). A comparison of the cell phone driver and the drunk driver. *Human Factors, 48*(2), 381-391.

Strayer, D. L., Drews, F. A., & Johnston, W. A. (2003). Cell phone-induced failures of visual attention during simulated driving. *Journal of Experimental Psychology: Applied, 9*(1), 23-32.

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 Washington, D.C.: AAA Foundation for Traffic Safety.

Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology. (1996). Heart rate variability: standards of measurement, physiological interpretation, and clinical use. *European Heart Journal, 17*, 354-381.

Tison, J., Chaudhary, N., & Cosgrove, L. (2011). National phone survey on distracted driving attitudes and behaviors. (Report No. DOT HS 811 555). Washington, DC: National Highway Traffic Safety Administration.

Victor, T., Bärgman, J., Boda, C., Dozza, M., Engstroem, J., Flannagan, C., et al. (2014). Analysis of Naturalistic Driving Study Data: Safer Glances, Driver Inattention, and Crash Risk: Safer.

Yang, Y., Reimer, B., Mehler, B., & Dobres, J. (2013). A field study assessing driving performance, visual attention, heart rate and subjective ratings in response to two types of cogntive workload. *Proceedings of the 7th International Driving Symposium on Human Factors in Driver Assessment, Training, and Vehicle Design*, Bolton Landing, New York, 397-403.

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.