

Evaluation of Two Types of In-Vehicle Music Retrieval and Navigation Systems

Jingyi Zhang^a, Avinoam Borowsky^{a*}, Bent Schmidt-Nielsen^b, Bret Harsham^b, Garrett Weinberg^c, Matthew R.E. Romoser^a, Donald L. Fisher^a

^a *Department of Mechanical & Industrial Engineering, University of Massachusetts Amherst, Amherst, MA*

^b *Mitsubishi Electric Research Laboratories, Cambridge, MA*

^c *Nuance Communications, Burlington, MA*

** Corresponding Author*

It has been documented that use of an In-Vehicle Information System [IVIS] can impair driving performance. In general, voice-controlled IVISs are considered less distracting than those controlled by touch and are therefore considered as less interfering with the driving task. However, certain types of voice-controlled IVISs may be better (less distracting) than others. In this study, we evaluated two variants of a voice-controlled in-vehicle music retrieval and navigation system in order to investigate the effect of a common affordance where the IVIS controls the pace and timing of a multi-turn interaction. Participants were asked to drive various scenarios in a simulator while engaging with either a user-paced IVIS or a system-paced IVIS. Driving performance measures and IVIS usability measures were collected. Our results indicated that when the driving task imposes higher mental effort (e.g., on a curved roadway), the user-paced IVIS was found to be more effective and safer to use, whereas when the driving task imposes low mental effort (e.g., on a straight roadway), the user-paced and the system-paced IVIS was less effective but still safer.

INTRODUCTION

Recent evidence shows that one of the primary causes for drivers' distraction is attributed to engagement in a secondary task, contributing to about 22% of all crashes and near-crashes (Klauer et al., 2006). As In-Vehicle Information Systems (IVISs) proliferate and become more complicated to operate, there is a major concern that this will be associated with an increase in the long glance durations inside the vehicle which are known to greatly inflate crash risk (e.g., Wierwille and Tijerina, 1998; Horrey and Wickens, 2007).

Studies have shown that interactions with IVISs while driving negatively affect drivers' performance. Chisholm et al. (2008) and Garay-Vega et al. (2010), for example, have shown that interactions with touch-controlled IVISs impaired drivers' ability to anticipate hazards and control their vehicles, as well as increased their Perception Response Time (PRT) to hazardous situations. The authors suggested that such driving-related deficiencies were the result of being engaged with in-vehicle systems that required multiple interactions to complete a secondary task while driving. They argued that such systems imposed demands on both the motor and visual systems, as well as attentional resources (Garay-Vega et al., 2010) – the resources that are critical for safe driving.

Voice-controlled IVISs, on the other hand, are considered less distracting than those controlled by manual input and therefore less likely to compete with the primary driving task. This is consistent with the multiple resource theory of attention (Wickens, 2002), suggesting that there are different pools of resources that can be utilized in parallel. Tasks that require the same input (visual or auditory), or output (manual or speech) modalities will interfere. Interacting with a touch-controlled IVIS, for example, requires visual inputs and manual outputs, which are exactly the same inputs and outputs required by the driving task. There will be no doubt that such two tasks will interfere with each other. Voice-controlled IVISs might have less impact on the driving task, as they require different input and output modalities from those required by the driving task (Ranney et al., 2005; Maciej et al., 2009).

Yet others have argued that even drivers who use voice-controlled IVIS are not immune to distraction. Several studies have demonstrated that listening to a radio or speaking while driving may deleteriously affect driving. Lee et al. (2001), for example, investigated drivers' ability to follow a lead vehicle while operating a speech-based email system in a driving simulator. Occasionally, the lead vehicle suddenly braked and drivers had to respond as soon as possible. Compared with the baseline group, drivers who engaged in the secondary task at the moment that the lead vehicle braked were 30% slower to respond. Drivers subjectively estimated that the workload was higher when the secondary task was included. A follow-up study done by Jameson et al. (2004) has confirmed the results. Both of the studies suggest that drivers' attentional resources have limited capacity for carrying out multiple tasks at one time regardless of the modality's type (Muller et al., 2011), which is to say, distracted drivers are more likely to be involved in a crash no matter what.

Clearly, further investigation is warranted into if certain types of voice-controlled interfaces are better than others. In this study, a driving simulator was used to compare and assess two variants of a voice-controlled, commercial IVIS. One, which is more commonly used, is the "system-paced" IVIS. A system-paced IVIS is designed so that the user initiates a voice-controlled task (usually by pressing a push-to-talk button), and then the timing of all further interactions in that task are controlled by the system. In particular, the system prompts the user for each subsequent utterance. The user does not have control over the timing of his/her responses to the system.

The other type of the IVIS is the "user-paced" IVIS. In this system the user must take an action to notify the system when speech is about to begin (usually by pressing the push-to-talk button before beginning to speak). The user has control over the timing of each utterance, but must perform an additional manual action for each utterance.

The system-paced design is motivated by a desire to remove the manual action associated with each spoken response subsequent to the initial request to activate the system. Such a design is expected to reduce the motor-related resources need-

ed to use this system. However, there is a tradeoff: the user-paced design allows the user to choose when to speak, so the users may choose *not* to speak at moments when the driving task imposes a higher mental effort and to postpone interactions with the system until the moment when mental resources are available. Thus, it is expected that the user-paced IVIS should be safer to use in either situation due to its self-pacing feature. Regarding to the usability factor, a user-paced IVIS would be less effective on the roads which require more attention and higher mental effort, but would be as effective on the straight roads, representing low mental effort, since users can choose to be engaged more often with the secondary task. The hypotheses are summarized in Table 1.

Measurement	Driving Condition	
	High mental effort (Curved road)	Low mental effort (Straight road)
Safety: Mean, StDev of -Velocity -Headway -Lane position	User-paced better than System-paced	User-paced better than System-paced
Usability -Total no. of errors -% of tasks completed	User-paced worse than System-paced	User-paced equal to the System-paced

Table 1 Summary of hypotheses

METHOD

Participants

Thirty-six participants (10 females and 26 males) took part in the experiment as paid volunteers. Participants' age varied from 18 to 30 years old, with an average age of 21.6 (SD=3.1) years and an average domestic driving experience of 3.7 (1.9) years. Individuals were students and staff recruited from the University of Massachusetts Amherst and the surrounding community. Participants received monetary compensation up to \$25 for their one-hour participation. All participants had normal or corrected to normal vision, normal contrast sensitivity and normal color vision.

Apparatus and Stimuli

Voice-Controlled Retrieval and Navigation Systems: control, interface, and tasks. The Mitsubishi Electric Research Laboratories (MERL) developed two variants of a prototype voice-controlled IVIS which allowed the user to access various types of information that would be commonly found in an IVIS: Music, Navigation (address information), and Contact information (phone number). Specifically, the two systems were the *system-paced* and the *user-paced*. The database of both systems was identical. The controls and display were configured as follows. (1) A steering wheel-mounted control panel contained a back button and a push-to-talk button, which were the manual inputs for the system. The back button was used to go back to the previous system state when participants entered an incorrect option (users could also correct an input error by returning to the initial dialog state using a voice command "Main Menu"). (2) A single dashboard-mounted 7" LCD was used to display the menu options. The average verti-

cal visual angle between the participants' direct forward gaze and the in-vehicle LCD was approx. 30 degrees.

The IVIS interface is typical of commercial IVISs in that it has a hierarchical (tree) menu structure with a voice dialog state for each node in the tree. The initial dialog is the root of the tree and can be accessed at any time by the voice command "Main Menu". At the Main Menu, there are options for each of the kinds of information available: "navigation", "music" and "contacts" (Fig. 1a). Each of these options or submenus contains other submenus, and so on (Fig. 1b). The IVIS is state-dependent in that only a small set of speech commands are allowed in any particular dialog state. In general, the available commands are displayed on the screen and when the state is entered, the user is prompted with state specific instruction, usually an audio list of the available commands. Best matches for the requested command are displayed as a list according to the order of relevance on the screen.

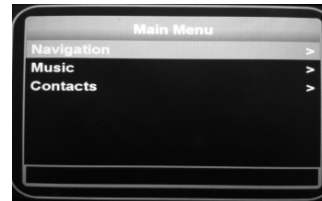


Fig 1a) Main Menu

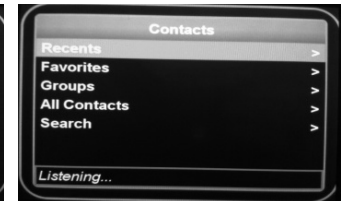


Fig 1b) Submenu of Option "Contacts"

Participants were asked to use one of the two variants of the IVIS while driving to accomplish three types of retrieval tasks corresponding to the kinds of information contained in the IVIS database: (1) Navigation, (2) Music, and (3) Contacts. The "Navigation" tasks started with an instruction asking the driver to navigate to a specific place or navigate to a given address. These general instructions were followed by more specific instructions such as "show the address on the map", "set this address as a destination", "add it to favorites", or "call the phone number." Similarly, Contacts tasks began with an instruction to find a particular contact name followed by a more specific instruction such as "send text message", "call phone number", "add to favorites", or "show home address on map". Music tasks were somewhat different from the others in that users were asked either to find a specified song, a specific song from a specified album, or to find a song by a specified artist. No further actions were required. Participants in both the system-paced group and the user-paced group were given the same tasks.

System-paced Voice-Controlled Retrieval and Navigation System. As mentioned above, in the system-paced IVIS, the actions required to perform a task were paced by the system: the user presses the push-to-talk button to begin the initial dialog and then all following conversational turns are initiated by the system by prompting the user with audio instructions. The system-paced IVIS cannot be interrupted while the audio prompt is given. A press on the push-to-talk button that occurs while the audio prompt is being played will be ignored.

User-paced Voice-Controlled Retrieval and Navigation System. Users of the "user-paced" IVIS must press the push-to-talk button every time they wish to speak to the system. The user-paced IVIS allows the user to interrupt the audio prompt at any time by pushing the push-to-talk button. This terminates the prompt so that the user may begin speaking immediately.

Eye Tracker. The Applied Science Laboratories (ASL) Mobile Eye, a lightweight tether-less Eye Tracking System (ETS), was used to monitor participants' eye movements. The system uses pupil-corneal reflections as the measurement principle. The sampling and output rate is 30 Hz and the system allows the driver's head a full range of motion. The system's accuracy is 0.5 degrees of visual angle.

Driving Simulator. The fixed-based driving simulator in the Arbella Insurance Human Performance Laboratory (HPL) at the University of Massachusetts Amherst was used in this experiment. The simulator, manufactured by Realtime Technologies Inc., consists of a full size 1995 Saturn sedan. The virtual driving scene is displayed on three screens which subtend 150 degrees of visual angle horizontally and 30 degrees vertically. The images were displayed at a resolution of 1024 x 768 pixels in each screen with a refresh rate of 60 Hz. A surround sound audio system is embedded in the simulator.

Virtual Drives. Participants were asked to drive three virtual drives. A drive is defined as a specific route that a participant needs to navigate from start to finish. The three virtual drives designed for this study included two drives (Drive 1 and 3) in an urban environment (Fig. 2) and one drive (Drive 2) in a suburban environment (Fig. 3). Two scenarios were embedded in Drive 2. They were selected to help evaluate the effects of the different variants of the IVIS. A scenario was defined as a specific segment of a drive during which an event would be triggered. Drives 1 and 3 were used as fillers while Drive 2 was used to evaluate the two variants of the IVIS.



Fig 2 City Street



Fig 3 Suburban Road

Experimental Design

Participants were pseudo-randomly assigned to three groups of equal size: the system-paced group, the user-paced group, and the control group. The order of the presentation of the three drives was counterbalanced across participants. During each scenario, drivers in the user- and the system-paced groups were asked to engage with a secondary IVIS task. The purpose of the secondary task was to evaluate driving performance as well as the usability with respect to the two variants of the IVISs in various driving conditions.

Scenario descriptions and measures. Drive 2 contained two scenarios embedded in an alternating sequence of straight and curved sections of roadway in a suburban environment. Participants started driving on the straight road, and ended driving on the "S"-type curved road. There was a lead vehicle in front of the participant's vehicle throughout the drive. The speed of the lead vehicle varied from 15 mph to 55 mph. Participants were asked to maintain a 2-3 seconds gap between their vehicle and the lead vehicle. Drive 2 was designed with repetitions of sections such that the secondary task (interacting

with IVIS) only appeared in the second straight and second "S"-shaped curved section (the two scenarios). The total number of interactions with the system during the task, the task completion status (i.e., if participants finished the task or not), the total number of participants' errors (an interaction was considered an error if, for example, a participant forgot to push the button while using the user-paced system, or a participant said the instruction before the system was in a receiving mode while using the system-paced system) and the vehicle simulator data were monitored.

Procedure

Participants completed a practice drive and once familiarized with the simulator they received instructions on how to operate their in-vehicle system (either the system- or user-paced system) and had time to practice it until they felt comfortable with its operation. Participants assigned to the control group were not involved in any procedure with respect to the systems. After practice, an eye calibration procedure was conducted and each participant completed the three drives. At the end of the driving section participants were asked to complete a NASA TLX Workload questionnaire..

RESULTS

As aforementioned, the secondary tasks in Drive 2 were initiated only on the second repetition of the straight and curved sections. If not specified otherwise, the variables listed below were analyzed using a Linear Mixed Model with a random intercept in SPSS. Two fixed effects were included in the initial model: (1) road geometry (straight road vs. "S"-type curved road) and (2) system type (system- vs. user-paced). Participants were included as a random effect. A backward elimination procedure (beginning with a model including all variables and then taking out non-significant effects) was applied afterwards to explore the effect of the different independent variables.

IVIS Usability (System-Related Measures)

In order to fairly evaluate the usability of the two variants of the IVIS, the total number of participant errors made while interacting with the system, the total number of interactions with the system during the scenarios, and the task completion rate were analyzed.

Total number of errors. An analysis was undertaken of the total number of errors participants made while engaging with the system. A participant error was defined as any kind of invalid interaction with the system. For example, when interacting with the user-paced system, an error would be recorded if a participant forgot to push the button when he or she was about to interact with the system; or when interacting with the system-paced system, an error would be recorded if a participant began interacting with the system when the system was not in a recording mode.

In order to normalize the dependent variable, the total number of interactions with the system was included as a covariate in the initial model. There were two significant main

effects in the final model: (1) road geometry [F(1,41.64) = 10.47, p<.01] and (2) total number of interactions [F(1,41.71) = 21.17, p<.01], and two significant second order interactions: (1) system type and road geometry [F(1,21.26) = 14.23, p<.01] and (2) road geometry and total number of interactions [F(1,41.90) = 9.22, p<.01]. The mean of the number of errors made in the curved section (Mean=.64, SE = .17) was more than double the number made in the straight section (.25, .19). Figure 4 shows the estimated means (average number of errors per participant) and the standard errors of the interaction between road geometry and system type.

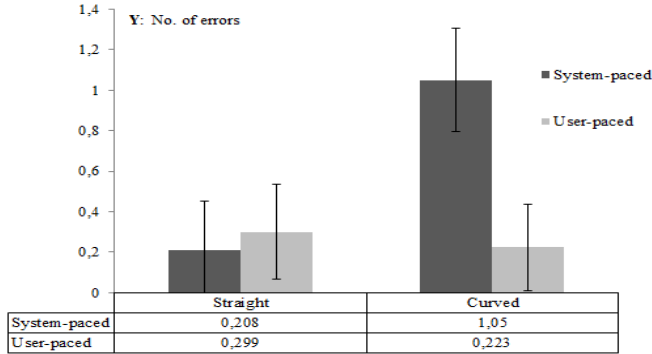


Fig 4 Interaction between road geometry & system type – estimated mean number of errors per participant

A LMM analysis of the two different road geometries respectively was undertaken to further explore the nature of the second order interaction. No significance of system type was found in the straight section, however, there was a significance of the system type [F(1, 21)= 8.955, p<.01] in the curved section, suggesting that the system-paced users produced more errors on the curved road than did the user-paced users (Fig. 4).

Task completion rate. Task success rate in each system was analyzed to determine which type of system was more accessible. The dependent variable is whether or not a participant completed the task (1- completed in the given time; 0 - otherwise). Although the distance for each participant to complete the task was the same, the time might slightly vary due to the various longitudinal velocities. A logistic regression with a random intercept in the framework of the Generalized Estimating Equations (GEE) was conducted. The fixed effects were system type and road geometry; participants were included as a random effect. A similar backwards elimination procedure revealed one significant main effect of road geometry [$\chi^2_{1,wald} = 5.75$, p = .016] and one marginally significant interaction between road geometry and system type [$\chi^2_{1,wald} = 2.95$, p = .086]. The estimated probability of completing the task in the straight section (Mean = .45, SE = .11) was almost three times larger than this probability in the curved road section (.12, .07). Figure 5 below depicts the marginally significant interaction.

A pairwise comparisons post hoc analysis using a sequential Bonferroni correction was applied to examine the interaction in detail. The task completion rate for the system-paced users was higher in the straight section than in the curved section, while no significant differences emerged in the user-paced group. Within the straight section, the task success rate

of the system-paced users was higher than the user-paced users; no difference appeared within the curved section.

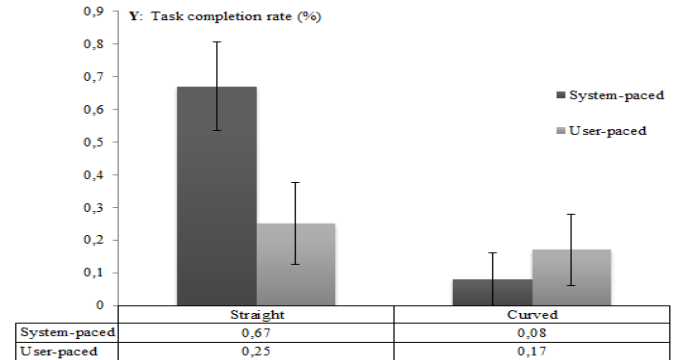


Fig 5 Interaction between road geometry & system type – task completion rate

Safety (Driving-Related Measures)

The average speed and headway, the standard deviation of speed and headway, and the standard deviation of lateral lane position were considered as criteria to evaluate the safety of the IVISs. Headway (distance in meters) could only be calculated on the straight section due to system restrictions. The longitudinal velocity and the lateral lane position were analyzed for both road geometries.

Average and standard deviation of longitudinal velocity. One significant main effect of road geometry [F(1, 66)= 1464.867, p<.01] was found after conducting the analysis. On the straight roadway, participants tended to drive faster on average (Mean = 34.21 unit: meters/sec, SE = 1.08) than on the curved roadway (20.10, 1.08). There was a non-significant interaction between the two fixed effects: a marked trend (p=.12) showed that on average, the user-paced group drove slower (Mean=33.32, SE =1.14) than both the system-paced group (34.34, 1.14) and the control group (34.97, 1.14) in the straight section, whereas no difference among the three groups appeared in the curved section.

With respect to the standard deviation of longitudinal velocity, there was a significant main effect of road geometry [F(1,66)= 26.82, p<.01]. The standard deviation of the longitudinal velocity was larger in the straight section (Mean=3.72, SE=.15) than in the curved section (2.60, .15). A marked trend (p=.187) indicated that in the curved section, the standard deviation of the longitudinal velocity of the system-paced group (Mean = 2.98 unit: meters/sec, SE = .27) was larger than that of the user-paced group (2.55, .27). The control group was somewhere between the two other groups (2.26, .27). No difference was found among the groups on the straight roadway.

Average and standard deviation of headway. The LMM included only one fixed effect of system type due to the fact that the headway could be calculated only in the straight section. There was a significant main effect of system type [F(2,28)= 6.585, p<.01]. A pairwise comparison post hoc analysis, using the Bonferroni correction, revealed that the control group drove closer to the lead vehicle (Mean=30.45 unit: meters, SE=3.76), than both the system-paced group (47.56, 3.76) and the user-paced group (47.31, 4.16). There was no significant difference between the system-paced group and the user-paced group.

No significant difference was found with respect to the standard deviation of headway on the straight section.

Standard deviation of lateral lane position. A log normal transformation was conducted to normalize the variable before carrying out the LLM analysis. A significant main effect of road geometry [F(1,68) = 344.71, p<.01] and a marginally significant main effect of system type [F(2,68) = 2.452, p<.1] were found. The road geometry analysis revealed that drivers had a larger standard deviation in the curved section [.58, CI = 1.19] than in the straight section [.15, .31]. A pairwise comparison post hoc analysis on the system type using the Bonferroni correction revealed that the user-paced group [.26, .58] had a significantly (marginal difference) smaller standard deviation (p<0.108) than the system-paced group [.32, .70]. The control group [.30, .66] was not different from the other two groups. Table 3 summarizes the results from the statistical analyses for safety and usability measures.

Measurement		Driving Condition	
		High mental effort (Curved road)	Low mental effort (Straight road)
Usability	Total number of errors	User < Sys	NS*
	Task completion rate	NS	User < Sys
Safety	Longi. Velocity	Average StDev ⁺	User < Sys
	Headway	Average StDev	N/A
	Lateral Position	StDev	User < Sys (on curved and straight sections altogether)

Table 3 Summary of statistical analysis for safety and usability measures
 *: User – User-paced IVIS; Sys – System-paced IVIS; NS – none significance
 †: StDev – Standard Deviation

DISCUSSION

The present study evaluated two variants of a voice-controlled in-vehicle information system in terms of usability (number of tasks completed, number of participant errors), as well as their effect on driving safety (the various dependent measures discussed above). Although engaging with a secondary task while driving will almost always interfere with the driving task, it was interesting to test whether a well-designed interface may dampen this interference and improve the usability and safety of the system. The evaluated system variants were: (1) the user-paced IVIS and (2) the system-paced IVIS.

The study results have partially confirmed our hypotheses (Table 1). Regardless of which system being used, participants made fewer errors and were more likely to complete a task in the straight section than in the curved section.

In the curved section, where more cognitive attention was required, the user-paced users made fewer errors than the system-paced users and drove more safely (the standard deviation of the longitudinal velocity and lane deviation were smaller, Table 3). Such results underscore the fact that the user-paced IVIS was safer and more effective to use when the cognitive demand is high. Although we expected that the user-paced IVIS would be safer we did not expect it to be more effective. One explanation for this unexpected finding is that the number of secondary tasks completed by drivers in the curved section

was very low with either IVIS being used, suggesting that all drivers decided to ignore the secondary task. But due to the nature of the system-paced IVIS, the interferences with the driving task requires more cognitive attention.

In the straight section where less cognitive attention was required, the system-paced users completed more tasks than the user-paced system users. Such a result was against our previous hypothesis: the user-paced IVIS would be as effective as the system-paced IVIS. One reasonable explanation would be the design of the user-paced IVIS required drivers' input to initiate each utterance and forced their eyes off the forward roadway more frequently, so that the users felt unsafe to do so. Meanwhile the system-paced IVIS kept prompting for instructions and forced the users to interact with. And this is exactly why, as previously expected, the system-paced IVIS was less safe to use (Table 3). It truly compromised the safety.

Most of the in-vehicle glances from the participants using the user-paced system occurred while looking for the push-to-talk button. It would be interesting to perform a further study to see if the usability of the user-paced IVIS would be improved if the control panel for both systems could be refined.

ACKNOWLEDGEMENT

Portions of this research were funded by grants from the Mitsubishi Electric Research Labs, the National Institutes of Health (1R01HD057153-01), and the National Science Foundation (Equipment Grant SBR 9413733 for the partial acquisition of the driving simulator).

REFERENCES

- Chisholm, S.L., Caird, J.K., & Lockhart J. (2008). The effects of practice with MP3 players on driving performance. *Accident Analysis and Prevention*. Vol. 40, No. 2, p. 704-713.
- Garay-Vega, L., Pradhan, A.K., Weinberg, G., Schmidt-Nielsen, B., Harsham, B., Shen, Y., Divekar, G., Romoser, M.R.E., Knodler, M., & Fisher, D.L. (2010). Evaluation of different speech and touch interfaces to in-vehicle music retrieval systems. *Accident Analysis and Prevention*. Vol. 42, No. 3, p. 913-920.
- Horrey, W.J., & Wickens, C.D. (2007). In-vehicle glance duration: distributions, tails, and model of crash risk. *Transportation Research Record*. No. 2018, p. 22-28.
- Jameson, A.H., Westerman, S.J., Hockey, G.R., & Carsten, M.J., (2004). Speech-based e-mail and driver behavior: effects of an in-vehicle message system interface. *Human Factors*. Vol. 46, No. 4, p. 625-639.
- Klauer, S.G. (2006). The impact of driver inattention on near-crash/crash risk: an analysis using the 100-car naturalistic driving study data. *National Highway Traffic Safety Administration*.
- Lee, J.D., Caven, B., Haake, S., & Brown, T.L., (2001). Speech-based interactions with in-vehicle computers. The effect of speech-based e-mail on drivers' attention to the roadway. *Human Factors*. Vol. 43, p. 631-640.
- Maciej, J., & Vollrath, M. (2009). Comparison of manual vs. speech-based interaction with in-vehicle information systems. *Accident Analysis and Prevention*. Vol. 41, No. 5, p. 924-930.
- Muller, C., Weinberg, G., & Vetro, A. (2011). Multimodal input in the car, today and tomorrow. *IEEE Multimedia*. Vol. 18, No. 1, p. 98-103.
- 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*. Vol. 47, No. 2, p. 439-454.
- Wickens, C.D. (2002). Multiple resources and performance prediction. *Theoretical Issues in Ergonomic Sciences*. Vol. 3, No. 2, p. 159-177.
- Wierwille, W.W., & Tijerina, L. (1998). Modeling the relationship between driver in-vehicle visual demands and accident occurrence. *Vision in Vehicles VI, Elsevier Science, Amsterdam, Netherlands*. P. 233-243.