Exploring In-Vehicle Systems

Input and Output Modalities in Safety Critical Use Contexts

Brit Susan Jensen and Nissanthen Thiruravichandran, Aalborg University 2009

Title: Exploring In-Vehicle Systems – Input and Output Modalities in Safety Critical Use Contexts.

Project period: Spring 2009

Project group: D610B

Brit Susan Jensen

Nissanthen Thiruravichandran

Supervisor: Mikael B. Skov

Table of Contents

Introduction

Human computer interaction (HCI) is the study of the interaction between human and computer. HCI is an extensive, multi-disciplinary field, which comprise computer science, cognitive science and psychology. Understanding the interaction between human and computer is an essential aspect of system development and has expanded concurrently with advances in technology.

We are students specialising in human computer interaction, our preceding aim is to design, evaluate and implement interactive systems in accordance with requisites of a given context.

Advancements in communication and information systems have introduced novel means to present, retrieve and apply information; such innovations may introduce changes in user perception and behaviour in a given context. We believe that these changes are important to address and hereby elaborate on how information systems set perceptual and behavioural changes in motion; through this knowledge we may be able to support the promotion of user satisfaction and safety. Furthermore, when exploring the interaction between human and computer it is crucial to consider the related contextual factors, since each context involve specific requisites. For example, the use of a handheld computer in a static context, such as an office, allows the user to attentively interact with the computer. A dynamic context on the other hand, where the user is on the move, necessitates attention to the surroundings while interacting with the computer. When introducing a device from a static context to a dynamic context, the behaviour originally associated to the device changes and new behavioural patterns emerge.

The emergence of mobility

Over the years, technology has facilitated practices, which were traditionally common in the home and office spaces, to be performed on the go. Mobility is becoming an essential part of our lives. Devices are becoming more compact, whilst an extensive amount of functionality is being incorporated. Users in dynamic contexts are given access to functionalities that were originally associated to stationary use contexts – for example telecommunication and access to the internet.

Car companies, hardware manufacturers and software developers have assessed that implementing extraneous functionalities in vehicles is profitable due to the number of potential drivers. Since then, the market of in-vehicle systems has been growing exponentially over the last three decades [3] – a proliferation, which has been set in motion by consumer demands, a reduction in hardware costs and significant innovations in communication and information technology. This development is not expected to come to a halt. Today, in-vehicle systems comprise a multitude of systems – for example; music players, car computers, DVD players, automotive navigation systems and collision detection systems. The common denominator of these systems is the increased safety risk they can impose by diverting the driver's attention from the primary task of driving.

In-vehicle systems often have their origins in settings, which are less safety-critical, which makes it questionable to which extent their usage and design is suitable in a vehicle-centric context. Research on traffic accidents show that the use of in-vehicle systems is involved in traffic accidents [2, 5], which could indicate that these systems, in their current state, have not been fully optimized to the conditions of their new setting. Several technologies have been developed and implemented into cars in an attempt to alleviate safety risks caused by in-vehicle system interaction – For example, remote controls are included when purchasing entertainment systems, head-up displays are implemented in the front windscreen to present important information to the driver. An example of a more complex safety system is the workload management system [5], which attempts to determine when the driver is preoccupied with more essential tasks. The system changes the availability of certain functionalities and pieces of information in accordance with the context.

We find this tendency to be troublesome, due to the fact that an interaction problem imposed by an in-vehicle system is sought to be solved by implementing additional systems in the vehicle. A different approach to address the problem of in-vehicle induced accidents needs to be taken. We find it important to improve the interaction between driver and in-vehicle system as opposed to implement intermediate safety systems to alleviate the problem.

Attention and distraction

When researching the field of car safety and the usage of secondary in-vehicle systems, it is also crucial to touch upon the concepts of attention and distraction.

Attention can be defined as the ability to concentrate and selectively focus or shift focus between selected stimuli [8]. Hamilton describes this in short as; *'consciousness voluntarily applied, under its law or limitations, to some determinate object'* [6]. The object would, within our selected field, be the different stimulus associated with and necessary to, the act of driving a car. Lansdown amongst others [2, 4, 5, 7], refers to this act as the driver's primary task, which in the case of driving would involve operating the car and navigating it safely.

In this context it is also relevant to acknowledge the limitations in relation to attention. A general term for these limitations within the field of vehicle research is distraction. Green describes distraction as anything that grabs and retains the driver's attention, diverting focus away from the primary task of driving the car [5]. Distraction is in this connection a broad concept [1, 2], which can involve eating while driving, internal thought processes, secondary systems, other passengers in the vehicle and the surroundings. A specific focus within the field of in-vehicle research, concerning distraction and attention, is the dynamics between the primary task of driving the car and secondary tasks, which involves use of in-vehicle system.

Since many in-vehicle systems are not crucial or relevant to the safety of the driver, it would seem that the degradation of secondary task systems could be accepted or this reduced performance simply should be taken into account when these systems are developed. The problem with this viewpoint is the fact that it shifts the responsibility of allocating attention solely on to the driver. This is problematic when researchers like Green [4] state that drivers will go to great lengths to complete a secondary task and rarely abandon a task once it has been initiated. With a critical primary task, like driving, this illogical behaviour and distribution of attention between the primary and secondary task can in worst case scenarios endanger the safety of the driver and the surroundings. This tendency stresses the importance of focusing on how people actually behave when using in-vehicle system.

Research focus

Previous research state that in-vehicle systems can cause drivers to divert their attention from the primary task of driving and hereby induce safety risks. Further research on how in-vehicle systems affect drivers is imperative in order to identify adequate interaction techniques and ways to present information. In our effort to contribute to the field of in-vehicle systems, we strive to answer the following research questions,

- *How do different configurations of input modalities affect driving performance?*
- *How do different configurations of output modalities affect driving performance?*

The first question concerns input techniques for interacting with secondary tasks and their effect on drivers, whereas the second question focuses on presentation and retrieval of information in relation to in-vehicle systems. The research questions served as the focus for two papers. In the first paper we sought to address both questions by comparing various configurations of an in-vehicle system with an equal emphasis on input and output modalities in order to explore how these modalities separately affect driving performance. The first paper inspired us to focus solely on output; hence the second paper addressed the second research question. We compared three output configurations of a stand-alone GPS system to assess their influence on driving performance.

Research Paper 1

In-vehicle interaction – The Separate Effects of Input and Output

Previous research on in-vehicle systems have identified that in-vehicle systems can cause driver distractions, which may induce severe safety risks. Studies which focus on alleviating driver distraction tends to focus on creating new input modalities without acknowledging output as an independent component. This is a troublesome tendency and in our study, we sought to put equal emphasis on both input and output in order to examine their separate effects on driving performance and eye glance behaviour. This research paper presents a comparison of four different combinations of input and output modalities. The purpose of this was to create an adequate combination of systems configurations, which would enable us to assess how they separately affect drivers. In this research paper we attempt to answer the following questions,

- *How do different configurations of input modalities affect driving performance?*
- *How do different configurations of output modalities affect driving performance?*

To answer this, we developed a music player, which consisted of four different configurations of input and output; gesture-visual, gesture-audio, touch-visual and touch-audio. To evaluate the configurations an experiment was conducted in a HCI laboratory, where we created a medium-fidelity driving simulator for the evaluation. Thirty-two participants were distributed equally amongst the four configurations and asked to complete a series of tasks while driving the simulator. We recorded and analysed the following; driving performance, eye glance behaviour and ability to successfully interact with the system in order to complete the assigned tasks.

Through the experiment we uncovered a number of interesting findings. Our results show that out of the four different configurations – the gesture-audio configuration had by far the lowest number of glances, but also more longitudinal control errors and longer task completion times. In a broader view we found that gesture input resulted in significantly less eye glances compared to touch output, but also decreased driving performance and increased task completion times. While audio output caused more longitudinal control errors and longer task completion time compared to visual output – visual output accounted for considerably more eye glances and interaction errors. Through the results we found no indications that fewer eye glances entail increased driving performance. Decreased driving performance could however be related to the presence of audio output, which would imply that audio output increases cognitive workload.

Research Paper 2

GPS Navigation – Evaluating Output Modalities for Safety Critical Contexts

In-vehicle systems provide compelling means to enhance mobility and serve a variety of purposes, which involve navigational assistance, communicative support, entertainment et cetera. Important research contributions have been made to the field of GPS systems. Nonetheless, further elaboration on how the navigational output affects driving performance is needed - especially when considering that previous research on other in-vehicle systems highlight how secondary tasks can lead to accidents. This study presents a comparison of three different output configurations of a GPS system. We aim to answer the following question,

- *How do different configurations of output modalities affect driving performance?*

In order to shed light on this matter, we conducted a field experiment in actual traffic. Three different system configurations of a stand-alone GPS system (audio, visual and audio-visual) were evaluated. The experiment involved 30 participants who were presented with four scenario-driven tasks that involved driving to predetermined locations. The participants were assigned to each configuration. We ensured than an equal distribution of GPS users and non-users was attained. We recorded driving performance and eye glance behaviour and we also collected qualitative data through observations and interviews.

We found that participants in the audio configuration performed best when assessing the measurement variables. Not surprisingly, we discovered a substantial amount of eye glances during field trials, which involved the visual configuration, but also a considerable number of incidents, where decreases in driving performance occurred. Participants in the visual configuration had significantly more speeding and lateral control violations compared to participants using the audio configuration. When using the audio-visual configuration resulted in a reduced glance frequency when compared to visual, but did however not improve driving performance. Interestingly, when assessing the primary driving task performance variable we found no significant difference between the visual and the audio-visual configurations, even though there are significant differences in eye glance behaviour. Although the audio configuration proved to be the most favoured when considering driving performance, the user satisfaction inquiries show a preference for having both output modalities available.

Conclusion

In this master's thesis we set out to explore how input and output of in-vehicle systems affect driving performance. Our contribution to the field of in-vehicle systems constitutes two research papers. In the first paper we sought to answer how different configurations of input and output affected driving performance. In the second paper we evaluated three output configurations of a GPS system, in order to shed light on how output affects driving performance. Our overall research questions were as follows,

- *How do different configurations of input modalities affect driving performance?*
- *How do different configurations of output modalities affect driving performance?*

In our initial study we aimed to compare different configurations of an in-vehicle system with an equal emphasis on both input and output modalities, in order to examine how they separately affect driving performance. Our study shows that when designing for in-vehicle systems it is important to consider the separate effects of input and output modalities.

The use of gesture input lead to significantly fewer eye glances in comparison to touch input, nonetheless gesture input still caused decreased primary driving task performance and longer task completion times. Audio output resulted in more longitudinal control errors as well as significantly longer task completion times when compared to visual output. Visual output, on the other hand, accounted for significantly more interaction errors and a substantially higher number of eye glances. Looking at the individual input and output configurations, our results show that gesture-audio by far has the lowest number of eye glance occurrences, but it also resulted in longer task completion times and more longitudinal control errors compared to the other configurations. Furthermore we were able to confirm that glances could be attributed to input techniques, as the users had to visually obtain the position of the system in order to interact with it. We did however not find a relation between the amount of eye glances and errors relation to the primary driving task performance.

In our second study we strived to clarify how different configurations of output modalities affect driving. We compared three output configurations of a stand-alone GPS system; audio, visual and audio-visual combined. The GPS system is a highly output oriented device, which made it an adequate platform for evaluating output modalities. Thirty participants attended our field experiment, which was conducted in real traffic.

Our results show that visual output not only causes a substantial amount of eye glances, but also results in decreased driving performance. While the introduction of audio output in combination with visual output reduces the frequency of glances, we found this to have no effect on driving performance. This indicates that the presence of audio output may induce additional cognitive workload, nevertheless audio output is beneficial when taking eye glance behaviour and glance tendencies into consideration. Even though the audio configuration proved to be the most favourable in terms of road safety, the user satisfaction inquiries show a preference for having both output modalities available.

In the first study we did not identify a relation between the amount of eye glances and errors related to the primary driving task performance. We did however see a pattern in the second study where the presence of visual output affects primary driving task performance significantly. An interesting result, which relates to the relation between driving performance and glances, is the fact that we did not find any differences in driving performance between audio-visual participants and visual participants, even though there is a significant difference in glance frequency. This could raise the question if reducing the number of glances by adding auditory output, is as beneficial as one could initially think. Research has shown that auditory output causes cognitive workload [10] and our results for the audio-visual and visual configurations seem to indicate that the increase in cognitive workload is just as devastating as the additional glances.

References

1. Bach, K. M., Jæger, M. G., Thomassen, N. G., In-Vehicle Systems Research: A Classification of Approaches to Understand, Measure and Evaluate Attention. Department of Computer Science, Aalborg University 2007.
2. Brooks, C. and Rakotonirainy, A. 2007. In-vehicle technologies, Advanced Driver Assistance Systems and driver distraction: Research challenges. I.J. Faulks, M. Regan, M. Stevenson, J. Brown, A. Porter and J.D. Irwin (Eds.). Distracted driving. Sydney, NSW: Australasian College of Road Safety. 471-486.
3. Broy, M. 2006. Challenges in Automotive Software Engineering in ICSE, ACM Press.
4. Chewar, C.M.,McCrickard, D.S., Ndiwalana, A., North, C., Pryor, J. and Tessendorf, D., Secondary task display attributes: optimizing visualizations for cognitive task suitability and interference avoidance.in Proceedings of the symposium on Data Visualisation (VisSym ’02), Eurographics Association 2002, pp. 165 – 171.
5. Green, Paul. Driver Distraction, Telematics Design, and Workload Managers: Safety Issues and Solutions. University of Michigan Transportaion Research Institute. SAE International (2004)-21-0022.
6. Sir W. Hamilton Dictionary.comhttp://dictionary.reference.com/browse/attention
7. Lansdown, T. C., Brook-Carter, N., Kersloot T., (2004). Distraction from Multiple In-Vehicle Secondary Tasks: Vehicle Performance and Mental Workload Implications in Ergonomics, Vol. 47, No. 1, Taylor & Francis Group, 91–104.
8. MedTerms.com http://www.medicinenet.com/hemin_for_injection/glossary.htm
9. Christiansen, L. H., Frederiksen, Y. N., and Ranch, A. 2009. The Invasion of Information systems: All Your Bases Are Belong to Us, Department of Computer Science, Aalborg University.
10. Strayer, D. L., Drews, F. A., (2007) Cell-Phone-Induced Driver Distraction. University of Utah.

Appendix A, Research Paper 1.

Appendix B, Research Paper 2.

Appendix C, Synopsis.

In this master's thesis we set out to explore how input and output of in-vehicle systems affect driving performance. Our contribution to the field of in-vehicle systems constitutes two research papers.

The market of in-vehicle systems has grown exponentially over the last three decades – a proliferation, which was set in motion by a reduction in hardware costs as well as innovation in communication and information technology. In-vehicle systems often have their origins in settings, which are less safety-critical, which makes it questionable to which extent their usage and design is suitable in a vehicle-centric use context. Several studies on in-vehicle systems have been conducted and it has generally been found that in-vehicle systems can cause driver distraction, which may induce severe safety risks. We find it important to elaborate on the dynamics between the primary task of driving the car and secondary tasks, which involves use of in-vehicle system.

Today, in-vehicle systems comprise a multitude of systems – for example; music players, car computers, DVD players, automotive navigation systems and collision detection systems. Further research on how in-vehicle systems affect driver behaviour is imperative in order to identify adequate interaction techniques and ways to present information.

In the first research paper we assessed various configurations of an in-vehicle system with an equal emphasis on input and output modalities – through this assessment we sought to explore how these modalities separately affect driving performance. A touch screen based music player that comprised of four different combinations of input and output modalities was developed for our experiment; audio-gesture, audio-touch, visual-gesture and visual-touch. In order to assess these combinations of input and output, we conducted an experiment in a medium-fidelity driving simulator. Thirty-two (16 male and 16 female) participants attended the experiment. We assigned an equal number of male and female participants to each of the configurations. The participants were asked to complete 32 tasks that involved input and output. When analysing the data we applied the following measurement variables: primary driving task performance (lateral and longitudinal control errors), secondary driving task performance (interaction errors and task completion time) and eye glance behaviour (glances below .5 seconds, between .5 and 2.0 seconds, above 2.0 seconds).

The results of our evaluation show that when addressing in-vehicle systems design, separating input and output modalities has an impact. The use of gesture input resulted in significantly fewer eye glances in comparison to touch input, nevertheless gesture inputs still resulted in inferior primary driving task performance and longer task completion times. Audio output caused the test subjects to commit more longitudinal control errors in addition to significantly longer task completion times compared to visual output. Visual output, on the other hand, accounted for significantly more interaction errors and a considerably higher number of eye glances. Looking at the individual input-output configurations, our results show that gesture-audio by far has the lowest number of glances, nonetheless also longer task completion times and more longitudinal control errors compared to the other configurations.

Based on the first paper we decided to explore output modalities further. We evaluated how different output modalities affect driving performance by comparing three output configurations of a GPS system; audio, visual and audio-visual combined. We conducted a field experiment in real traffic; we hereby sought to elicit insight into user behaviour within an approximated natural context. Thirty participants (15 GPS users and 15 non-users) attended the experiment. The participants were randomly assigned to a configuration until an equal distribution of GPS system users and non-users was attained – we assigned five GPS system users and five non-users to each of the three configurations (which constitute three groups of ten). All field trials started at the Computer Science Department at Aalborg University. The field trials comprised of four scenario-driven tasks – the tasks involved driving to predetermined locations and collecting associates of the University. The participants were not given tasks during driving – each task was presented prior to the associated driving segment. The segments consisted of both rural and densely populated areas in order to expose the participants to varied traffic environments and areas of Aalborg. We applied the following measurement variables for the data analysis: primary driving task performance (lateral and longitudinal control errors), secondary driving task performance (navigational errors and task completion time) and eye glance behaviour (glances below .5 seconds, between .5 and 2.0 seconds, above 2.0 seconds).

Our results indicate that visual output not only causes a substantial amount of eye glances, but also leads to a considerable decrease in driving performance. While the introduction of audio output in combination with visual output reduces the frequency of glances, we interestingly enough found this to have no effect on driving performance. This indicates that the presence of audio output may induce additional cognitive workload, nonetheless audio output is beneficial when considering eye glance behaviour and glance tendencies. Although the audio-only configuration proved to be the most favourable in relation to driving performance, the user satisfaction inquiries show a preference for having both output modalities available.

Based on the results of the two studies, we find that both a separate and cohesive assessment of input and output modalities are important in order to understand in-vehicle interaction and its effects on driving performance.

In-vehicle Interaction – The Separate Effects of Input and Output

Brit Susan Jensen, Nissanthen Thiruravichandran
Department of Computer Science, Aalborg University
Selma Lagerlöfs Vej 300, 9220 Aalborg East, Denmark
{raifire, nisunivu}@cs.aau.dk

PREFACE

This study was conducted in collaboration with Lars Holm Christiansen, Nikolaj Yde Frederiksen and Alex Ranch from the Department of Computer Science, Aalborg University during 9th semester. We have slightly modified Related Work, Results and Discussion. We have partly re-written; Abstract, Experiment and Limitations. The Introduction has been completely re-written.

ABSTRACT

New precautionary design measures are imperative in order to address driver distraction that may be attributable to additional information components in vehicles. Research within the field of in-vehicle systems tends to focus on input modalities without acknowledging output as an independent component. In our study, we sought to put equal emphasis on both input and output in order to examine their separate effects on driving performance and eye glance behaviour. We evaluated four combinations of input modalities (touch and gesture) and output modalities (visual and audio) in a driving simulator. Our results showed that gesture input resulted in significantly fewer eye glances when compared to touch input, but induced worse primary driving task performance. Audio output resulted in a substantially lower number of eye glances, but introduced significantly longer task completion times and low primary driving task performance when compared to visual output. The results emphasize that output influenced vehicle and system operationality, hence an equal emphasis on input and output modalities is essential to system design for safety-critical contexts.

Author Keywords

Gesture based interaction, touch interaction, attention, eye glances, auditory output, visual output, in-vehicle systems.

INTRODUCTION

An increasing amount of technology becomes integrated in vehicles, these technologies serve a number of different purposes ranging from entertainment to navigational support. Several studies on in-vehicle systems have been conducted and it has generally been found that in-vehicle systems can cause driver distraction, which may induce severe safety risks [4, 9].

The complexity of in-vehicle systems correlates, with the level of distraction the systems induce. This correlation between complexity and distraction instigates a precautionary approach to the development of in-vehicle systems in order to address driver distraction and the associated safety risks [2, 3]. Current studies, which address driver distraction and safety factors related to in-vehicle systems, have a tendency to primarily focus on input modalities as opposed to output.

Our aim in this paper is to assess various configurations of an in-vehicle system with an equal emphasis on input and output modalities – through this assessment we seek to explore how these modalities separately affect driving performance. A music player that comprised of four different combinations of input and output modalities was developed for our experiment. We applied a touch screen for the music player – the touch screen is an interaction technology commonly seen in a variety of in-vehicle systems. The flexibility in its application capabilities, low price, and utilization of a natural way of interaction, presumably make it an apparent choice for an in-vehicle system. The paper is structured as follows; we initially present previous research on in-vehicle systems, and secondly we describe the music player we developed for the experiment. Hereafter we outline the proceedings of the experiment and present the results. Finally, the results are discussed.

RELATED WORK

When researching the field of vehicle safety and the usage of in-vehicle systems, it is crucial to touch upon the concepts of attention and distraction. Attention can be defined as the ability to concentrate and selectively focus or shift focus between selected stimuli [13]. Hamilton describes attention as '*consciousness voluntarily applied, under its law or limitations, to some determinate object*' [10]. In the vehicle-centric environment, the determinate object would be vehicle operationality, which associates and necessitates a variety of stimuli. Lansdown, amongst others [2, 5, 9, 12], refers to this act as the primary task of the driver, which implicates monitoring the environment and executing manoeuvres. Disruption of attentive vehicle operationality is defined as distraction. Green describes distraction as anything that grabs and retains the attention of the driver, shifting focus away from the primary task of operating the vehicle [2, 4, 9].

An important aspect within the field of in-vehicle research concerning distraction and attention is the dynamics between the primary task of driving the car and secondary tasks, which comprise use of in-vehicle systems. In-vehicle systems often have their origins in different and less safety-critical use-contexts. One example is the car radio which emerged as home entertainment and later on was implemented into cars. When traversing two such different use-contexts it is important to consider how the original design and usage might affect driving performance and whether it is suitable as an in-vehicle system – such considerations are significant since research identifies the use of in-vehicle systems as a potential cause of traffic accidents [9, 4]. Green, amongst others, point out that most drivers will go to great lengths to complete a given secondary task and rarely abandon a task upon initiation [9]. With a critical primary task, like driving, this seemingly irrational distribution of attention between the primary and secondary task, could in worst case scenarios, endanger the safety of the driver and their surroundings. Lansdown et al. acknowledges this unsettling tendency in a study focusing on distraction imposed by in-vehicle secondary systems [12].

The identified tendencies within the area of traffic safety research have also inspired further studies to find new interaction techniques for in-vehicle systems with the intention to alleviate driver distraction when operating an in-vehicle system – thus implying the need to identify an interaction technique that surpasses the capabilities of the traditional tactile interface. Recent work on in-vehicle systems suggests an overall preference for systems utilizing gestural input.

In a comparative study Geiger et al. set out to evaluate the use of dynamic hand movements (gestures) in order to operate a secondary in-vehicle system and compare it to a traditional haptic (tactile) interface [8]. The following parameters were used for comparison; errors related to vehicle operationality, tactile-gesture recognition performance and the amount of time drivers did not have both their hands on the steering wheel. The experiment showed that use of the tactile interface resulted in high task completion times and the system had low recognition performance when compared to the gesture based interface. The gesture interface allowed users to perform the primary task appropriately, the users also found the gesture interface more pleasant and less distracting. A recent study by Alpern and Minardo supports the findings put forth by Geiger et al. In the study they evaluated gestures through an iterative development of an interface for control of secondary tasks [1]. In the final iteration of their experiment, they noted that users made fewer driving errors when compared to a traditional tactile interface. Findings from both studies indicate that gestures could be a viable option for secondary in-vehicle systems.

Bach et al sought to shed light on how perceptual and task-specific resources are allocated while operating audio systems in a vehicle-centric environment [2]. Three system configurations – a conventional tactile interface, a touch interface and an interface, which recognized gestures as input – were evaluated in two complementary experiments. The experiments suggest an overall preference for the gesture-based configuration, as it enabled users to allocate their attention to the lateral and longitudinal control tasks. The tactile configuration lacked intuitiveness; hence the system necessitated perceptual and task-specific resources in order to be operated, thus disrupting primary task performance. The touch configuration introduced a reduction of overall task completion time and interaction errors, when compared to the tactile and gesture interfaces.

While the future prospect of using gestures as an interaction method for in-vehicle systems seems promising, little consideration is given to the possible influence of output modalities. To address this it would be essential to separate input and output in order to clarify how different output modalities might affect interaction when combined with various input modalities. Bach et al acknowledges the lack of focus on output in relation to in-vehicle systems as a limitation in their comparative study. Their pre-eminent research focus was on system input as opposed to output, albeit the output modalities differed for each configuration. The variation in output modalities could have affected the findings – the results does not elicit, which output modality is suitable in a safety-critical use context – thus implying the necessity for an elaborate study on output modalities in order to elicit how output could exert influence upon primary and secondary task performance.

The aim of our study is to compare different configurations of an in-vehicle system with an equal emphasis on both input and output modalities, in order to examine how they separately affect driving performance. We aim to confine system variables (with regards to input and output), and hereby approximate a strict comparative study. We intend to accomplish this through a study of visual and auditory output in combination with either touch or gesture input. The rationale behind this combination is the duality in touch screens interaction abilities, which supports both touch and gesture interaction and the polarity in the two different sensory channels of the output.

IN-VEHICLE INTERACTION

In this study we developed an in-vehicle system – a music player. This choice was inspired by Bach et al amongst others [1, 2, 14]. The system comprise the following four configurations; touch input with visual output, touch input with audio output, gesture input with visual output and gesture input with audio output. The system is designed to fit an 8 inch touch sensitive screen, and the graphical user interface in all configurations is divided into the same output and input areas, to keep the interaction areas the same for all conditions. Furthermore, the output area of the screen is covered by a clear plastic shield to discourage deliberate input and prevent accidental input.

Input

We distinguish between two input modalities; conventional touch screen based input with graphical buttons, and gesture-based input using the touch screen as a drawing canvas (as seen in figure 1). The graphical layout of the two touch configurations is inspired by Bach et al [2] and our goal was to keep it as simple as possible, while still providing the necessary functionalities. To facilitate easy interpretation the icons on the buttons resemble icons commonly used on music players. Furthermore, the buttons are grouped according to their functionality. The layout includes a 'Song info' button, which is only enabled in the touch-audio configuration, but is included in the touch-visual configuration to ensure consistency in layout across all configurations. The size and spacing of the buttons is chosen based on previous research on touch screen buttons [6, 16, 17]. Input is only possible by pressing the buttons, which work according to the click-on-release principle. This means that the buttons are activated only when the finger has left the button.

The gesture-based systems have no buttons. Instead, the systems are controlled by gestures drawn directly on the screen using a finger. The gestures are inspired by Pirhonen et al [14] and facilitate the same functionality as touch input. The only gesture we have changed is the 'Song info' gesture, which is performed by drawing a line straight down followed by a line straight up, without the finger leaving the canvas. This was chosen to resemble the 'i', which denotes 'information'. The gestures can be executed anywhere within the input (grey) area of the screen except in the output (white) area.

Output

We implemented two modes of output; visual output using icons and text and audio output using ear-cons and voice. Visual and audio output is not used simultaneously at any point. We also distinguished between two kinds of output; feedback on user input and information on system status.

The visual feedback was implemented using visual cues to inform the user of the result of his or her actions. For the touch-visual system, this is done by changing the appearance of buttons to indicate they have been pressed (by inverting the colours). When the volume is all the way down, pressing the 'Volume down´ button will change its appearance to denote a disabled state (shown in a grey colour). The same principle applies to the 'Volume up'

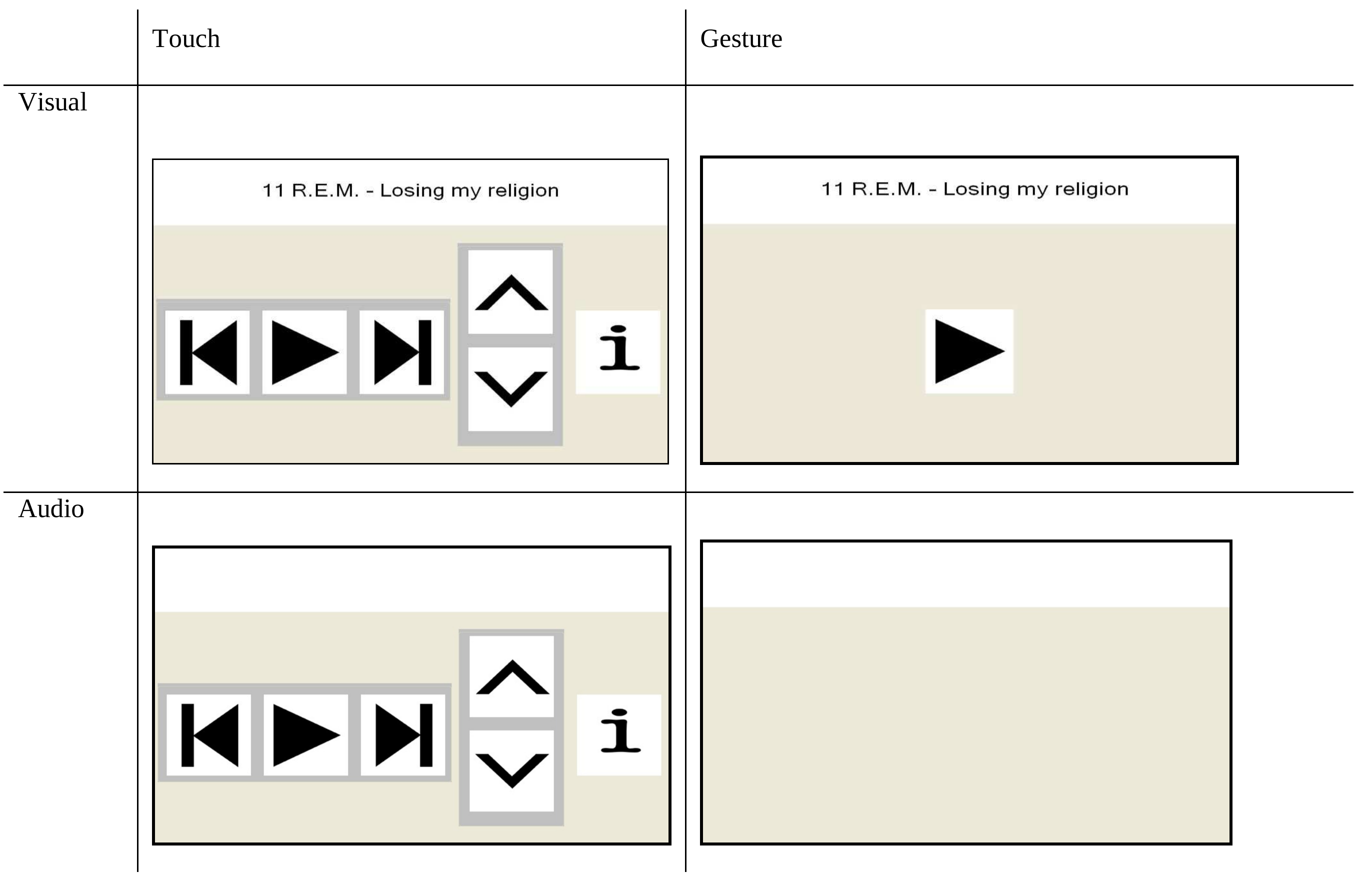


Figure 1. The graphical user interfaces for the four configurations. The top (white) part of the screen is reserved for output, while the grey area is for input. On the visual (top row) configurations the buttons are, from left to right, 'Next song', 'Play/pause', 'Previous song', 'Volume up' and 'Volume down', 'Song info'. In the figure for gesture/visual, the user has just performed the 'Play' gesture, causing the system to flash the 'Play' icon.

button. For the gesture-visual system, the same icons are used to indicate a recognized gesture. The icon corresponding to the recognized gesture will be displayed in the middle of the input area for a few seconds (as shown on Figure 1)

Audio feedback was implemented using earcons. In the touch-audio and gesture-audio systems, when the user either pushes a button or performs a gesture, the system will provide feedback in the form of a clearly audible 'click' sound. Following the same principle that applies to visual feedback, any attempt to adjust the volume either up or down when it is fully up or down, will result in a 'dong' earcon.

Output about the state of the system consists of information regarding the current song; the song's number in relation to the playlist, the artist and the title of the song. Visual output about the state of the system is provided by text in the output area of the screen and is available at all times. The audio output is implemented using playback of voice recordings containing the same information. These recordings are played by pushing the 'Song info' button or performing the 'Song info' gesture.

EXPERIMENT

The purpose of the experiment was to compare the four different configurations of the music player, in order to assess how the input and output modalities separately affected driving performance. In the following we will describe the proceedings of the experiment.

Experimental Design

We used a between-subject experimental design, where the independent variables comprise configuration (touch-visual, touch-audio, gesture-visual, gesture-audio) and the dependent variables comprise primary driving task performance (longitudinal control, lateral control), secondary driving task performance (interaction errors, task completion time) and eye glance behaviour.

		Input	
		Touch (N=16)	Gesture (N = 16)
Out put	Visual (N = 16)	N = 8	N = 8
	Audio (N = 16)	N = 8	N = 8

Figure 3. Experimental design

Participants

Thirty-two participants (16 male and 16 female) attended the experiment. All the participants carried valid driver licenses and had so for 0.5 – 29 years (M = 9.4, SD = 8.7) and drove by their own estimates between 100 – 30.000 kilometres per year (M = 6114.69, SD = 7987.9). The participants were aged between 21 and 56 years (M = 28.2, SD = 9.2). The average amount of kilometres driven per year, were the same for each participant group.

Setting

Our experiment was conducted in the HCI laboratory at Aalborg University, where we created a medium-fidelity driving simulator (as seen in Figure 2). The simulator consisted of a PC running Test Drive Unlimited, two car seats, a projector and a force feedback steering wheel with a brake and an accelerator. The game featured regular roads with traffic and regulation signs. The system was mounted on the right side of the steering wheel. The setup also included two sets of speakers; a set of 4.1 surround sound speakers which played the sound effects featured in the game, and a set of 2.1 stereo speakers for music playback. In front of the participants the game was projected onto the wall (see Figure 2 middle).

Procedure

We assigned an equal number of male and female participants to each of the configurations – hence, each group had four male and four female participants. The participants were asked to solve 32 tasks – 16 of the tasks primarily focused on system input – for example; 'Please skip two tracks'. The other 16 were mainly output oriented – for example; 'Please tell us the name of the artist and the

Figure 2. Pictures of the driving simulator. Left: the test manager sitting in the passenger seat next to the participant. Middle: A view of the road in the simulator (bottom part of the view included the speedometer). Right: The steering wheel and music player

title of the song that is currently playing'. Furthermore, when creating the tasks we attempted to ensure that they did not favour any of the four configurations. The tasks were divided into two sections of 16 questions. In the first section, the test manager guided the participants through a predetermined route, by telling them when and where to turn. In the second section the participants were instructed to drive as they pleased, whilst keeping within the parameters they initially had been instructed to comply with. The participants were instructed to drive the car between 40 – 60 km/h except when executing manoeuvres, which required a decrease in speed.

The demographical data of all the participants was gathered prior to the experiment. The test manager then briefed the participants by reading a text aloud, which introduced them to the proceedings of the test session. They were also shown how to operate the music player in the particular configuration they were to use during the test session. They were subsequently asked to repeat the interaction, in order to ensure they had understood the given instructions. The participants took a test run in order to familiarize themselves with the driving simulator. After the introduction the test manager moved the car to a predetermined location, which was the same for all test sessions. The participants were then instructed to start driving. The test manager gave the participants directions as well as tasks. The participants were instructed to initiate each task when they felt ready to do so.

Data Analysis

In the analysis of the data we adapted the variables from Bach et al [2] using similar measurement variables for assessing how the different modalities affect driving performance;

- Primary driving task performance.
- Secondary driving task performance.
- Eye glance behaviour.

All five of the authors of this paper participated in the analysis process, where each test session was reviewed by three authors. The focus of the analysis was to measure performance in relation to the above measurement variables.

Errors in primary driving task were defined as lateral and longitudinal control errors. A lateral control error was defined as lane excursions where the participant failed to stay within the two lines denoting the right hand side lane. Longitudinal control errors were defined as failure to maintain a speed within the instructed range of 40 – 60 km/h. A longitudinal error was recorded each time the participants went above or below the speed range. Staying at a wrong speed for a period of time only counted as one error. Errors in secondary driving task were defined as interaction errors and task completion time.

Interaction errors were defined as attempts to interact with the system that did not have the effect towards completion of the task that the participants expected. Task completion time was measured from the time the participants started solving the task, defined by either moving their hand from the steering wheel, or moving their head/eye gaze towards the system, until the task was completed.

Eye glances comprise the following three categories;

- Category 1, 0.5 seconds and below.
- Category 2, between 0.5 and 2.0 seconds.
- Category 3, above 2.0 seconds.

In order to ensure the highest possible uniformity in the interpretations of the data, two of the sessions were analyzed by all the authors collectively. This presented us with an opportunity to discuss the various types of incidents in the data, and subsequently to compile a set of directions to be followed in the following individual analysis.

The data analysis was done by individually reviewing the videos whilst logging instances of the abovementioned incidents in a spreadsheet. This meant viewing the videos frame by frame in order to precisely determine the length of each eye glance. Eye glances with a duration of 13 frames or less were categorized as category 1 (0.5 corresponds to 13 frames). Eye glances with an interval between 14 – 50 frames were categorized as category 2, while eye glances with duration above 51 frames were categorized as category 3.

As mentioned all the video for all 32 participants were reviewed by three authors. The three lists were then compiled into one final list containing all the incidents. This was done by way of majority vote; if for instance, only one of the reviewers had recorded an incident at a specific time, which neither of the two other reviewers had recorded, the end result was no incident by two votes, and so forth. Incidents where all the reviewers unanimously agreed were recorded as such. In situations were no majority vote could be secured, the video recording was viewed once again in order to reach a final verdict. An inter-rater reliability test (weighted Fleiss's Kappa) of the data gave $K = 0.70$, which corresponds to a substantial agreement.

RESULTS

In this section we present the results of our data analysis. The results are presented in three sections; Primary Driving Task Performance, Secondary Driving Task Performance and Eye Glance Behaviour. In each section we first compare the results for the two input modalities ($N = 16$), then the two output modalities ($N = 16$) and finally all four configurations ($N = 8$). The results were subjected to either two-tailed unpaired Student's t-tests or one-way repeated-measures ANOVA tests, as well as Tukey's HSD post hoc tests where applicable. The results are presented in tables, where all statistically significant differences at the 95% level are highlighted.

Primary Driving Task Performance

The variables for measuring primary driving task performance included lateral control errors (lane excursions) and longitudinal control errors (speed increases and decreases). Across the 32 test sessions, we identified a total of 256 lane excursions and 511 incidents of speed increases and decreases.

	Input Touch (N = 16)	Input Gesture (N = 16)	Output Visual (N = 16)	Output Audio (N = 16)
Lane excursions	7.19 (4.79)	8.81 (7.13)	8.63 (6.64)	7.38 (5.5)
Speed increases	6.31 (3.07)	6.69 (6.05)	**4.56 (3.41) -**	**8.44 (5.15) +**
Speed decreases	8.31 (7.42)	10.63 (5.02)	9.38 (2.04)	9.56 (6.79)
Total speed deviations	14.63 (8.50)	17.31 (7.67)	13.94 (5.78)	18.00 (9.63)

Figure 4. Means and standard deviations for primary task performance across the input and output modalities.

When comparing the primary task performance across the input modalities, we see no significant difference between any of the variables, although gesture input generally has a higher number of errors across all the variables (see Figure 4). Looking at primary task performance across the output modalities, reveals a significant difference in the number of speed increases, where visual has significantly fewer incidents than audio, $t = 2,04$, $p < 0.05$. However, there are no significant differences between each configuration when comparing the total number of speed deviations, although it is worth noting that the number of speed decreases and total speed deviations is higher for audio output than for visual output.

	Touch/Visual (N = 8)	Touch/Audio (N = 8)	Gesture/Visual (N = 8)	Gesture/Visual (N = 8)
Lane excursions	7.63 (4.87)	6.75 (5.01)	9.63 (8.28)	8.00 (6.23)
Speed increases	6.00 (3.66)	6.63 (2.56)	**3.13 (2.59) -**	**10.25 (6.54) +**
Speed decreases	6.38 (5.13)	10.25 (9.11)	12.38 (5.68)	8.88 (3.83)
Total speed deviations	12.38 (8.79)	16.88 (11.67)	15.50 (8.27)	19.13 (10.37)

Figure 5. Means and standard deviations for primary task performance across the four configurations.

If we look at the primary driving task performance variable and compare the four configurations (see Figure 5), we see a significant difference in the number of speed increases, $F(3, 28) = 3.95$, $p < 0.05$. A Tukey's HSD post hoc test revealed that there are significantly fewer speed increases in the gesture-visual configuration compared to the gesture-audio configuration ($p < 0.05$). A comparison of the remaining primary driving task performance variables shows no significant differences. But the results do show that the audio configurations have the highest number of total speed deviations.

Secondary Driving Task Performance

For secondary driving task performance we measured the total task completion time and identified a total of 1018 interaction errors. When comparing the input modalities, the results show only marginal differences in the number of interaction errors and the task completion time, although gesture does show a higher task completion time than touch, $t = 2.04$, $p < 0.19$ (see Figure 6).

	Input Touch (N = 16)	Input Gesture (N = 16)	Output Visual (N = 16)	Output Audio (N = 16)
Interaction errors	29.38 (19.69)	34.25 (29.99)	**40.69 (29.13) +**	**22.94 (16.82) -**
Task completion time	271.00 (62.13)	308.81 (95.20)	**256.94 (67.66) -**	**322.88 (82.40) +**

Figure 6. Means and standard deviations for secondary task performance across the input and output modalities.

Whereas the input modalities revealed no significant differences in secondary task performance, a comparison of the output modalities showed 77% more interaction errors for visual output compared to audio output. A t-test shows that this is a significant difference, $t = 2.04$, $p < 0.05$. The task completion times, however, were significantly longer for audio output, $t = 2.04$, $p < 0.05$.

	Touch/Visual (N = 8)	Touch/Audio (N = 8)	Gesture/Visual (N = 8)	Gesture/Visual (N = 8)
Interaction errors	42.38 (19.72)	16.38 (7.46)	39.00 (37.72)	29.50 (21.27)
Task completion time	**249.88 (24.28) -**	292.13 (81. 62)	264.00 (95.42)	**353.63 (75.66) +**

Figure 7. Means and standard deviation for secondary task performance across the four configurations.

Secondary driving task performance results revealed no significant differences in the number of interaction errors distributed amongst the four configurations (see Figure 7), even though the average number of interaction errors for the touch-audio configuration is less than half when compared to touch-visual and gesture-visual configurations, $F(3, 28) = 1.87$, $p < 0.16$. However, a significant difference does exist between the task completion times, $F(3, 28) = 3.06$, $p < 0.05$. A post hoc test showed that there is a significant difference between task completion times when comparing the touch-visual and gesture-audio configurations ($p < 0.05$).

Eye Glance Behaviour

We identified a total of 2371 eye glances divided into category 1 (below 0.5 seconds), category 2 (between 0.5 and 2.0 seconds) and category 3 (above 2.0 seconds) altogether. Of the total glances, around 60% occurred are related to touch input, which amounts to a significant difference when compared to gesture input, t = 2.04, $p < 0.05$. Looking at the individual eye glance categories, the results show a strong significant difference in the number of category 2 glances , where gesture input has substantially fewer glances, t = 2.04, $p < 0.01$. But when looking at the two other glance categories, touch has the fewest, although the difference is not significant (see Figure 8).

	Input Touch (N = 16)	Input Gesture (N = 16)	Output Visual (N = 16)	Output Audio (N = 16)
Category 1 < 0.5 s.	16.44 (13.85)	20.44 (12.09)	15.94 (11.85)	20.94 (13.88)
Category 2 0.5-2 s.	**71.88 (19.35) +**	**36.19 (36.66) -**	**76.06 (24.34) +**	**32.00 (13.88) -**
Category 3 > 2.0 s.	0.88 (1.36)	2.38 (3.74)	**3.19 (3.43) +**	**0.06 (0.25) -**
Total glances	**89.19 (19.10) +**	**59.00 (46.83) -**	**95.19 (30.14) +**	**53.00 (34.43) -**

Figure 8. Means and standard deviations for eye glance categories across input and output modalities.

Visual output accounts for 1523 (64%) of the total number of glances across output modalities, which amounts to an extreme significant difference, t = 2.04, $p < 0.001$. There is also an extremely significant difference in the number of category 2 glances with audio being significantly lower than visual, t = 2.04, $p < 0.001$. When considering the number of category 3 glances there is a strong significant difference when comparing the output modalities, where visual output yet again accounts for the majority of glances (visual accounts for 51 glances as opposed to audio, which accounts for 1 glance), t = 2.04, $p < 0.01$. On the other hand, audio output has marginally more category 1 glances (below 0.5 seconds) than visual output.

	Touch/Visual (N = 8)	Touch/Audio (N = 8)	Gesture/Visual (N = 8)	Gesture/Visual (N = 8)
Category 1 < 0.5 s.	8.88 (4.19)	24.00 (16.20)	23.00 (13.02)	17.88 (11.34)
Category 2 0.5-2 s.	**86.50 (12.40) + (+)**	**57.25 (12.62) + (-)**	**65.63 (29.44) +**	**6.75 (5.70) -**
Category 3 > 2.0 s.	**1.75 (1.49) -**	**0.00 (0.00) -**	**4.63 (4.27) +**	**0.13 (0.35) -**
Total glances	**97.13 (18.08) +**	**81.25 (28.83) +**	**93.25 (46.73) +**	**24.75 (17.40) -**

Figure 9. Means and standard deviations for eye glance categories across the four configurations.

Across the four configurations (see Figure 9), the touch-visual configuration accounted for 32% of the total number of glances, where 27% involved touch-audio, and while gesture-visual accounted for 31%, whereas gesture-audio only amounted to 8%.

A one-way repeated-measures ANOVA showed this difference to be extreme significant, $F(3, 28) = 13.59$, $p < 0.001$. Looking at the numbers, it is not surprising that the post hoc test revealed that the number of glances for the gesture-audio configuration was significantly lower than for any of the other configurations, $p < 0.01$.

Although touch-visual has substantially fewer category 1 glances when compared to touch-audio for instance, this does not represent a significant difference, but a one-way repeated-measures ANOVA indicates that it is close to significant, $F(3, 28) = 2.65$, $p < 0.07$. When assessing category 2 glances, however, an extremely significant difference exists, $F(3, 28) = 30.22$, $p < 0.001$. A post hoc test showed that gesture-audio has significantly fewer glances in this category than any of the other configurations, $p < 0.01$. This is perhaps not surprising, as gesture-audio accounts for just 8% of all the glances in this category. The post hoc test also revealed a significant difference between the number of category 2 glances (between 0.5 and 2.0 seconds) when comparing touch-visual and touch-audio, $p < 0.05$. When looking at category 3 glances (above 2.0 seconds), our results show an extremely significant difference in the number of glances, $F(3, 28) = 7.20$, $p < 0.001$. According to the post hoc test, gesture-visual has significantly more glances in this category than any of the other configurations, with $p < 0.01$ when compared to touch-audio (0 glances) and gesture-audio (1 glance), and $p < 0.05$ when compared to touch-visual.

DISCUSSION

Current conventional interaction techniques for in-vehicle systems have been linked to a decrease in driving performance. We set out to research how different input and output modalities affect driving performance in order to identify potential interaction techniques for safety-critical contexts. In the following we discuss and reflect on our results.

Separating input from output

Bach et al state that they are unsure what effect it has when their interaction techniques differ both in input and output – further studies are needed to address this issue. This is what we have done in our work, where the results show that a distinction between input and output is indeed an important one to make. Our results show that there is a significant difference in the number of glances when comparing across output modalities. This seems to imply that when conducting experiments with in-vehicle systems an undivided focus on both input and output is important.

Input

In our experiment, touch interaction proved to perform faster and with less interaction errors, compared to gesture input, although this difference is not significant.

Our initial assumption was that touch input would require more eye glances than gesture input, since the participants presumably need to visually obtain the position of the buttons before commencing interaction. This is also supported by our findings were we find a strong significant difference in category 2 glances (between 0.5 and 2.0 seconds) and a significant difference in the total number of glances, which is in line with [1,14]. In fact, the touch technique accounted for 51% more glances than the gesture technique, with respect to the total amount of eye glances. This number is even greater when viewing the category 2 glances separately, where touch input accounts for almost twice as many glances (98%) as gesture input. This is in line with Alpern and Minardo's findings which state that gesture interfaces are not attention free, but help drivers solve their task while allowing them to keep their eyes of the road [1].

The difference in eye glance behaviour can perhaps in part be explained by the fundamental design of the systems. When interaction fails with a touch button based interface, or if several interactions have to be performed in succession, users might have a tendency to use more glances in order to ensure or reassure that the correct button is being pressed. Similarly one might suspect that with gesture input, the user only has to visually confirm the position of the system before being able to issue one or more commands without looking, as opposed to finding the correct button on the screen. This could partly explain the difference in the number of glances.

Before conducting the experiment we also had the assumption that gesture input would have relatively more category 1 glances (below 0.5 seconds) compared to touch, the rationale being that the aforementioned visual confirmation of the position of the system should not take long. However, none of our findings corroborate this assumption. In terms of the number of interaction errors, the two input techniques show no significant difference to each other. In line with to the findings of Bach et al [2] our results also show touch as the fastest of the two input forms, although not significantly.

Output

When looking at primary driving task performance there is some difference between audio and visual output. Only in the number of speed increases is this difference significant – in favour of visual output. However the total number of speed deviations is not significantly different. It is unclear what these results indicate since the number of speed decreases is almost identical, and the total amount of speed deviations shows no significant difference.

When comparing task completion time for the two output techniques of our system, there is a significant difference between the two, in favour of visual output. We believe this is due to the nature of the audio output technique. When using audio to solve tasks requiring output from the system, the user first has to hear, and then process the information they are provided with before being able to solve the task. With visual output the user only has to read the information in order to solve the task, which presumably takes less time. Perhaps the user has already seen the information while performing another task, which further decreases the time required to solve certain tasks with the visual output technique.

Another interesting finding is that there is a strong to extreme significant difference in the number of eye glances between visual and audio. We believe that there are several reasons for this difference: first and foremost, the nature of audio output gives less incentive for looking at the screen, since it does not contain any visual information, nor does it give any kind of visual feedback. Presumably, users of touch-audio have more motivation for looking at the screen, compared to gesture-audio, since they still need to locate the buttons on the screen. However, for both configurations it applies that when issuing commands to the system, nothing is gained from looking at the screen, since no feedback is presented there. This is evidently different from the configurations with visual feedback, where there is no way of obtaining feedback other than looking at the screen, which would explain the difference in the number of glances. To sum up, Audio output leads to a higher task completion time, but fewer eye glances compared to visual output. And, aside from a significant difference in the number of increases in speed, there is no overall significant difference in the primary driving task performance when comparing the configurations.

In terms of road safety it may be argued that an increase in task completion time is acceptable, in exchange for fewer eye glances, which would allow more attention to the road. Our results do not however show a correlation between the number of glances and primary driving performance, which is similar to the findings in Bach et al [2]. However other studies [8, 15] state that a correlation between eye glance behaviour and driving performance does exist. In line with Gellaty [7], it is not difficult to imagine that more visual attention on the road is preferable, since eye glances are arguably the driver's primary method of assessing danger signs in traffic. However, further studies are required in order to determine whether this is really the case. This is also indicated in a study on how hands-free mobile phone conversations affect driving performance [18]. Strayer and Drews 2007 [18] state that even if drivers conversing through a hands-free mobile phone direct their gaze at the road, they often fail to notice objects in the driving environment, since their attention is occupied by the mobile phone conversation. However, the results in [18] relate to mobile phone conversations, which they claim differ qualitatively from other auditory tasks.

Although our results show that systems with audio output lead to distinctly fewer eye glances than systems with visual output, the results also seem to indicate that audio output comes at a price – namely an apparent drop in primary driving task performance. For instance, the number of speed increases and total number of speed deviations are higher for audio output than for visual output, although not significantly. This could indicate that listening to audio output while performing the driving task causes an increase in the cognitive load of the driver, thereby drawing mental resources away from the task of driving. This would be in line with a recent study in the field of brain research, which showed that driving while comprehending language – for example listening to voice messages from a hands-free mobile phone – results in a deterioration of driving performance [11]. Cognitive workload is also touched upon in Bach et al [2] in relation to their gesture-audio system. The system setup did not allow them to see an explicit connection between driving performance and output modality, which led them to attribute low driving performance to memory load (remembering gestures and the state of system). Another possible contributor to increased, or perhaps misaligned cognitive load, is the amount of the time the driver spends on solving a specific secondary driving task. As previously mentioned, our results showed that the subjects receiving audio output spent significantly more time on completing the tasks. Audio output might result in fewer glances, nonetheless the driver may still be cognitively occupied with a given task.

LIMITATIONS

Some of our test participants expressed difficulties with relating the simulated driving with a real-life traffic context, thus it is debatable whether testing within an authentic context, would have led to different results. A number of the participants mentioned the lack of tire noise and the fact that they were unable to orientate themselves through the side window, in addition to the sensation of movement, as some of the issues, which affected the notion of authenticity and therefore also their driving performance. This was partially because these factors provide the drivers with a sensation of movement, which facilitate an easier estimation of speed, when not looking at the road ahead. This could imply that simulated driving had an influence on the number of longitudinal control errors, this view is also supported by Bach et al. [2].

Our choice of case system represents a possible source of inaccuracy. The nature of the music player means that it will always provide a form of audio feedback, regardless of the chosen output modality. For example, pushing the 'Play' button will cause music to be played; increasing the volume will cause the music to become louder, et cetera. This means indicates that participants using visual output would not necessarily need to look at the screen to receive feedback.

CONCLUSION

As an increasing number of systems are making their way into cars, and pre-existing in-vehicle systems are becoming more advanced, research is needed to in order to elucidate how to design interaction techniques that consider the unique characteristics and requirements of the vehicular domain. Previous research has a tendency to focus primarily on the input aspect of in-vehicle interaction. The aim of this paper was to address this issue by placing equal emphasis on input as well as output, in order to investigate their separate effects on driving performance and eye glance behaviour. This was undertaken done through the evaluation of four different combinations of input and output techniques.

The results of our evaluation show that when addressing in-vehicle systems design, separating input and output modalities has an impact. The use of gesture input resulted in significantly fewer eye glances in comparison to touch input, nevertheless gesture inputs still resulted in inferior primary driving task performance and longer task completion times. Audio output caused the test subjects to commit more longitudinal control errors in addition to significantly longer task completion times compared to visual output. Visual output, on the other hand, accounted for significantly more interaction errors and a considerably higher number of eye glances. Looking at the individual input-output configurations, our results show that gesture-audio by far has the lowest number of glances, yet it also caused longer task completion times and more longitudinal control errors when compared to the other configurations.

Our results did not indicate that fewer eye glances necessarily entail better primary driving task performance. On the contrary, audio output, which has the fewest eye glances by far, seems to cause worse primary driving performance as well as longer total task completion times compared to visual output. This could imply that audio output has an effect on the mental workload of the driver, distracting their cognitive attention from the primary task of driving the car. Further research might shed more light on this phenomenon.

REFERENCES

1. Alpern, M and Minardo, K. 2003. Developing a Car Gesture Interface for Use as a Secondary Task, CHI 2003: New Horizons. Human-Computer Interaction Institute (HCII), Carnegie Mellon University.
2. Bach, K. M., Jæger, M., Skov, M. B., G., Thomassen, N. G. You Can Touch, but You Can't Look: Interacting with In-Vehicle Systems. Proceeding of the twenty-sixth annual SIGCHI conference on Human factors in computing systems. ACM 2008.
3. Bach, K. M., Jæger, M. G. 2007. A Classification of In-Vehicle Systems Research: Understanding, Measuring and Evaluation Attention, HCI Lab Technical Report.
4. Brooks, C. and Rakotonirainy, A. 2007. In-vehicle technologies, Advanced Driver Assistance Systems and driver distraction: Research challenges. Distracted driving. NSW: Australasian College of Road Safety.
5. Chewar, C.M., McCrickard, D.S., Ndiwalana, A., North, C., Pryor, J. and Tessendorf, D., Secondary task display attributes: optimizing visualizations for cognitive task suitability and interference avoidance in Proceedings of the symposium on Data Visualisation (VisSym '02), Eurographics Association 2002.
6. Colle, H. A., Hiszem, K. J. 2004. Standing at a kiosk: Effects of key size and spacing on touch screen numeric

keypad performance and user preference, Wright State University.
7. Gellatly, A. 1997. The Use of Speech Recognition Technology in Automotive Applications. Faculty of the Virginia Polytechnic Institute and State.
8. Geiger, M., Zobl, M., Bengler, K., Lang, M., Intermodal differences in distraction effects while controlling automotive user interfaces, Usability Evaluation and Interface Design, HCI 2001, Lawrence Erlbaum Ass.
9. Green, Paul. Driver Distraction. 2002. Telematics Design, and Workload Managers: Safety Issues and Solutions. University of Michigan Transportaion Research Institute. SAE International.
10. Hamilton: Sir W. Hamilton Dictionary http://dictionary.reference.com/browse/attention
11. Just, M., Keller, T. A., Cynkar, J. 2008. A decrease in brain activation associated with driving when listening to someone speak. Carnegie Mellon University.
12. Lansdown, T., Brook-Carter, N. and Kersloot, T. 2004. Distraction from Multiple In-Vehicle Secondary Tasks: Vehicle Performance and Mental Workload Implications in Ergonomics, Vol. 47. Taylor & Francis Group.
13. Medicin: Medicine Net http://www.medicinenet.com/hemin_for_injection/glossary.htm
14. Pirhonen, A., Brewster, S. and Holguin, C. (2002). Gestural and Audio Metaphors as a Means of Control for Mobile Devices in CHI Letters, Vol. No. 4.
15. Rockwell, T. H. 1988. Spare visual capacity in driving – revisited: New empirical results for an old idea in Gale, A. G., et al. Vision in Vehicles II, Elsevier Science.
16. Sears, A. Improving Touchscreen Keyboards: Design issues and comparison with other devices, University of Maryland 1991.
17. Sears, A., Revis, D., Swatski, J. Crittenden, R, Shneiderman, B. Investigating Touchscreen Typing: The effect of keyboard size on typing speed, University of Maryland 1992.
18. Strayer, D. L., Drews, F. A. 2007. Cell-Phone-Induced Driver Distraction, University of Utah.

GPS Navigation – Evaluating Output Modalities for Safety Critical Contexts

Brit Susan Jensen & Nissanthen Thiruravichandran
Department of Computer Science, Aalborg University
Selma Lagerlöfs Vej 300, 9220 Aalborg East, Denmark
{raifire, nisunivu}@cs.aau.dk

ABSTRACT

In this paper we explored how different output configurations of a GPS affect driving performance. We conducted a field experiment in real traffic with GPS users and non-users. Our results indicate that visual output not only causes a substantial amount of eye glances, but also leads to a considerable decrease in driving performance. While the introduction of audio output in combination with visual output reduces the frequency of glances, we interestingly enough found this to have no effect on driving performance. Although the audio-only configuration proved to be the most ideal in relation to driving performance, the user satisfaction inquiries show a preference for having both audio and visual output modalities available.

Author Keywords

Output modalities, in-vehicle systems, GPS system, eye glances, driving, field experiment.

INTRODUCTION

The market of in-vehicle systems has grown exponentially over the last three decades – a proliferation, which was set in motion by a reduction in hardware costs as well as innovation in communication and information technology [2, 4]. In-vehicle systems serve a variety of purposes, which involve navigational assistance, communicative support, entertainment et cetera.

The advancement of in-vehicle systems has incited debate and inspired research on road safety. In-vehicle systems may provide compelling means to enhance mobility [4, 17], albeit research has shown that these systems may distract the driver and hereby divert focus from the primary task of driving, which could lead to accidents [2, 8, 9]. In-vehicle systems have become increasingly sophisticated, due to the incorporation of additional functionalities and novel interaction techniques. Such progression may present the driver with tasks that are unrelated to driving (secondary tasks), which require more attentive interaction due to a higher level of complexity.

The Global Positioning System (GPS) was amongst the top selling consumer technologies in 2008[1]. In the recent years the GPS has become a subject of research. Current research on GPS systems has shed light on the way in which the utilization of these navigation systems may alter driving practices and affect the way people understand the environment in which they traverse [16]. Studies have also addressed important usability aspects by evaluating the learnability and memorability of a GPS system, in order to identify problems associated with first-time and infrequent use [14]. Furthermore, studies have raised awareness of the way in which navigational information should be presented in order to enhance user comprehension and satisfaction [11, 17]. These studies acknowledge that new design measures are needed to address the demands put forth by users. Evidently, important research contributions have been made to the field of GPS systems. Nonetheless further elaboration on how the navigational output affects driving performance is needed [11, 9] – especially when considering that previous research on other in-vehicle systems highlight how secondary tasks can lead to accidents.

In this paper we evaluated how different output modalities affect driving performance by comparing three output configurations of an off-the-shelf GPS system; audio, visual and audio-visual combined. We conducted a field experiment in real traffic; we hereby sought to elicit insight into user behaviour within an approximated natural context. This choice is supported by Devonshire et al. [7] amongst others [11, 17], who state that natural settings can account for the effects of driver choices and perceived risk through immersion. We chose a GPS system for our experiment to further promote immersion as it is commonly used in vehicles, moreover it is also highly output oriented; hence it may be regarded as an adequate platform for our evaluation. The paper is structured as follows; we initially present previous research on in-vehicle systems. Secondly, we describe the proceedings of the field experiment and subsequently, we present the results from the experiment. Finally, we discuss the results and put forth potential design implications.

RELATED WORK

Horberry et al. [12] amongst others [9] confirm that the level of complexity of a given secondary task is correlated with the level of driver distraction. Green [9] describes two concepts of distraction; *eyes-off-the-road* and *mind-off-the-road*. Drivers are required to have their eyes on the road in order to ensure safe driving; hence secondary tasks, which

[1] The NPD Group, press release 13th January 2009.

require visual attention, can induce safety risks. *Mind-off-the-road* concerns instances, where the driver's mind is occupied by matters unrelated to driving.

Several researchers have sought to alleviate driver distraction through novel means of interaction to support secondary tasks. Geiger et al. [8] conducted a comparative study in which they evaluated the use of a tactile interface and a gesture-based interface for secondary tasks. Their findings showed that the use of the tactile interface entailed high task completion times and low recognition performance when compared to the gesture based interface. The gesture-based interface enabled users to perform the secondary tasks accurately, the users also perceived the gesture-based interface to be less distracting in comparison to the tactile interface. Bach et al. [2] compared three interaction techniques for an audio system – a conventional tactile interface, a touch interface and a gesture-based interface. The interaction techniques were compared in two complementary experiments. The findings show an inclination towards the gesture-based interface, since it entailed low visual demand. The interface allowed users to complete secondary tasks, with significantly fewer eye glances in comparison to the tactile and touch interfaces. The tactile interface lacked intuitiveness; hence the system demanded perceptual resources in order to be operated and hereby diverting attention from the primary task. The touch interface introduced a lower task completion time and fewer interaction errors, in comparison to the other two interfaces.

Research on interaction techniques has identified promising input modalities for handling secondary tasks. While novel input techniques are important contributions to the field of in-vehicle systems, it is also essential to evaluate the potentialities and limitations of output modalities in safety-critical use contexts. Bach et al. [2] amongst others [11, 17] acknowledge that there is a need for further elaboration on how output modalities affect driving performance, in order to identify potential output modalities for in-vehicle systems.

George et al. [11] evaluated four configurations of a GPS system – auditory, auditory with landmarks, visual and visual with landmarks – in a simulated setting. The aim of the study was to clarify how much attention the systems required and how the participants perceived the use of the systems. The participants watched a video recording of a driver's view and received route guidance information. They were asked to press a button when they saw the intersections described by the GPS system. The study showed that the auditory configuration demanded less attention and provided the lowest reaction time when compared to the visual configuration. When supplementing the audio and visual configurations with information on landmarks no penalties to driver attention incurred. The participants mainly favoured the configurations, which included landmarks as opposed to the ones without landmarks. They generally expressed a slight inclination towards visual output. A recent empirical study by McCrickard et al. [17] partly confirms the results put forth by George et al. In the study they investigated trade-offs involving information conveyance by evaluating four information modalities – audio, audio with overhead map, visual and visual with overhead map – in a driving simulator. Their results show that the visual modality with an overhead map resulted in the highest number of driving errors and highest reaction time compared to the other modalities, whereas participants expressed that the information provided by the audio-based modalities were the most difficult to comprehend. They acknowledge that a more immersive setting could provide further validation of the results. Both [11, 17] studies indicate that audio output is a promising modality in terms of safety, whereas an inclination towards visual output is expressed in relation to user preference and comprehension.

Inspired by previous research we evaluated the navigational output provided by the GPS system. The majority of research on in-vehicle systems has been conducted in controlled and simulated settings [1, 11, 17]. We chose to conduct the experiment in the field due to the inadequacy of immersive settings within the area of in-vehicle research.

EXPERIMENT

The purpose of the experiment was to identify how different output modalities affect drivers and driving performance. In the following section we describe the system used in the tests and the proceedings of the experiment.

Experimental Design

This study utilized a between-subject experimental design, where the independent variables constituted the configuration type (audio, visual, audio-visual) and the dependent variables were primary driving task performance (longitudinal control, lateral control, violations related to traffic lights and directional indicators), secondary driving task performance (navigational errors and task completion time) and eye glance behaviour (0.5 seconds or less, 0.5 – 2 seconds, 2 seconds and above).

System Description

The GPS system used in the experiment was the TOMTOM GO 930 [6], which is a state-of-the-art model that has received favourable reviews. We chose this model with the presumption that this would minimise interaction errors during the field trials, due to its acclaimed design. The GPS came pre-installed with maps and a Points of Interest database covering Europe. The main input interface is a 4.3 inches touch screen with a 480 x 272 pixel resolution for visual output and an internal speaker for auditory output. With test participants having Danish as their native language we selected the language setting to match this in both visual and audio output. The three individual configurations consist of different combinations of output modalities. The audio configuration consisted exclusively of the system's audio output, while the visual configuration consisted exclusively of the system's visual output. The audio-visual configuration comprised both output modalities.

Auditory output compose of navigational instructions presented through pre-recorded speech (hence no speech output for street names was available) in a female voice. The female voice was chosen since it is easier heard in noisy environments [18]. Each instruction included an estimated distance and a direction – for example *'after 200 meters, turn left'* – followed by a repetition of the direction. If there was a need to perform a sequence of turns (within 200 meters of the first turn) this would be included in the instructions – for example *'after 200 meters, turn right and then turn left'*. On longer stretches of road (over 500 meters) the system would add an additional reminder.

Visual output in the GPS system consisted of the 'driving view' as seen in picture 1. This screen consists of a 3D map which shows the current part of the route that the driver is traversing. The selected route is marked in red, manoeuvres are illustrated with green icons and the current position of the vehicle is shown with a blue arrow. The lower part of the screen shows the appertaining navigation instructions which include estimated distance, arrival time, signal strength et cetera.

Picture 1. Example of the GPS in 'driving view'. The map is illustrated in 3D. In the map are street names, POI icons and zoom options. Navigation and system information are shown on the lower part of the screen.

Participants

Thirty participants ranging between 21 – 38 years of age (M = 25.2, SD = 2.65) attended the experiment. All participants (7 women and 23 men) carried valid driver licenses and had so for 3 – 19 years (M = 6.85, SD = 2.71) and drove by their own estimates between 0 – 40.000 kilometres per year (M = 7598.33, SD = 8557.9). On the basis of self-assessment – 9 participants answered that they had poor knowledge of Aalborg and its environs, 15 answered basic knowledge, while 6 claimed that they had good knowledge of Aalborg and its environs. The average amount of kilometres driven per year, were the same for each participant group. Likewise, each participant group were equally acquainted with Aalborg.

Picture 2. Example of the second camcorders view angle, which shows the driving view and the GPS system.

Setting

We conducted a field experiment in real traffic. All participants drove vehicles (C-segment – small family cars) equipped with manual transmissions. The field trials were conducted during daytime and in good weather conditions. Through these choices of conduct we also sought to approximate consistency between the field trials and hereby ensure an adequate basis for comparison of results.

The GPS system was – in concordance with the enclosed instruction booklet – affixed on the lower centre of the windscreen for all configurations, which provided the participants using visual output modalities with an unobstructed view of the visual output. The internal speaker volume was set to 75% for all field trials, which involved audio output, none of the participants found it necessary to further adjust the volume. For the field trials in which the audio configuration was applied, the GPS system was slanted in a manner, which ensured that the participants were only able to receive auditory output. The field trials were filmed using two camcorders – one of the camcorders was mounted on the dashboard in order to capture eye glance behaviour. The second camcorder was affixed on the front passenger seat to record lateral and longitudinal control errors, and driver view (as seen through the front windscreen).

We refrained from asking questions during the field trials in order to further limit safety risks. Dialogue only took place in the event that the participants initiated a conversation. The test manager was sitting next to the attending participant and an observer was sitting in the back. The test manager was assigned to ensure that the field trials proceeded as intended and to answer the participants' questions. The task of the observer was to collect qualitative data, such as the participants' utterances during the field trails. Prior to the actual test sessions we conducted two pilot tests in order to ensure that all the equipment was functioning.

Tasks

The field trials comprised four scenario-driven tasks (see Appendix C) – the tasks involved driving to predetermined locations and collecting associates of the University – for example '*Collect Lisa Nielsen who lives on Poseidonvej 15, 9210*'. By applying scenario-driven tasks, we sought to promote a natural setting for the field trials. The GPS system served as an optional component, thus allowing participants to approach a given task unassisted.

Procedure

The participants were classified as one of the following user types – GPS system users or non-users – the decisive factor for this classification process was based on whether or not the participants had used GPS systems previously. The participants were randomly assigned to a configuration until an equal distribution of GPS system users (15) and non-users (15) was attained – we assigned five GPS system users and five non-users to each of the three configurations (which constitute three groups of ten). We ensured that each group had at least one female participant (audio = 3, visual = 3, audio-visual = 1).

We initially collected demographic data of the participants through an interview. Subsequently, the participants were introduced to the GPS system and the proceedings of the field trial, where after they were asked to sign a consent form. All field trials started at the Computer Science Department at Aalborg University. The participants were not given tasks during driving – each task was presented prior to the associated driving segment. The estimated length of the entire route was 16 kilometres – the segments comprised both rural and densely populated areas in order to expose the participants to varied traffic environments and areas of Aalborg, which they may either be familiar or unacquainted with. The permitted speed limit ranged from 30 – 80 kilometres per hour in the four driving segments – we eschewed motorways due to safety concerns.

The participants were debriefed as a concluding segment of the field trials. The participants were first asked to complete a post-task questionnaire (five point Likert scale) on the GPS system – for example, the participants were asked; *'How helpful did you find the instructions provided by the GPS system?'* We then conducted a semi-structured interview on the answers given in the post-task questionnaire in order to clarify the underlying rationale of the participants' answers.

Data analysis

The collected data consisted of 30 video recordings and supplementary qualitative data. One of the video recordings was omitted due to incomplete data caused by a technical error in the recording equipment. We initially reviewed three randomly chosen video recordings collaboratively in order to establish guidelines for the subsequent individual video reviews. We reviewed 16 video recordings individually – 10 of the video recordings were reviewed by both authors to ensure procedural consistency. The review process constituted 35 lists of incidents, which were compared and merged into one list. When disagreements occurred both authors reviewed the video recordings in order to determine whether the concerned incident was valid or not. An inter-rater reliability test (weighted Cohen's Kappa) of the data gave $\alpha = 0.75$, which corresponds to a substantial agreement.

We applied measurement variables, which collectively constitute an adequate basis for the assessment of the output modalities [1, 2]. The variables encompass;

- Primary driving task performance (lateral control errors, longitudinal control errors, violations related to traffic lights and directional indicators).
- Secondary driving task performance (navigational errors and task completion time).
- Eye glance behaviour (glance category 1, glance category 2 and glance category 3).

Primary driving task performance constitutes variables, which concerns adherence to traffic regulations. Lateral control errors denote loss of lateral vehicle control, which comprise incidents of lane excursions. Longitudinal control errors denote incidents in which problems related to the control of vehicle velocity occurred (speed maintenance). Longitudinal control comprises three categories (in accordance with Danish legislation);

- Speeding level 1, incidents in which participants exceed the prescribed speed limit by three kilometres per hour (implies a speeding fine)[2].
- Speeding level 2, incidents in which participants exceed the prescribed speed limit by 30 percent (implies endorsement of license – one penalty point).
- Speeding level 3, incidents in which participants exceed the prescribed speed limit by 60 percent (implies revokement of license).

[2] The regulation was established due to inaccuracy in the measuring equipment.

Analysis of video recordings and supplementary qualitative data were undertaken in order to identify lateral and longitudinal control errors. We also recorded incidents in which participants did not adhere to the caution and stop signals assigned by traffic lights in addition to incidents where participants failed to activate the directional indicators as required by Danish traffic regulations.

Secondary driving task performance denotes incidents in which participants diverged from the specified route due to misinterpretation of the navigational information provided by the GPS system. We also recorded the completion time for each of the four tasks.

Eye glance behaviour is an acknowledged indicator of how driver attention is allocated [2]. Eye glances comprise the following three categories;

- Glance category 1, 0.5 seconds or less (0.5 seconds corresponds to 13 frames).
- Glance category 2, 0.5 – 2.0 seconds (corresponds to 14 – 50 frames).
- Glance category 3, 2.0 seconds and above (2.0 seconds corresponds to 51 frames).

Eye glances were identified by reviewing the video recordings frame by frame (the video recordings have a frame rate of 25 frames per second).

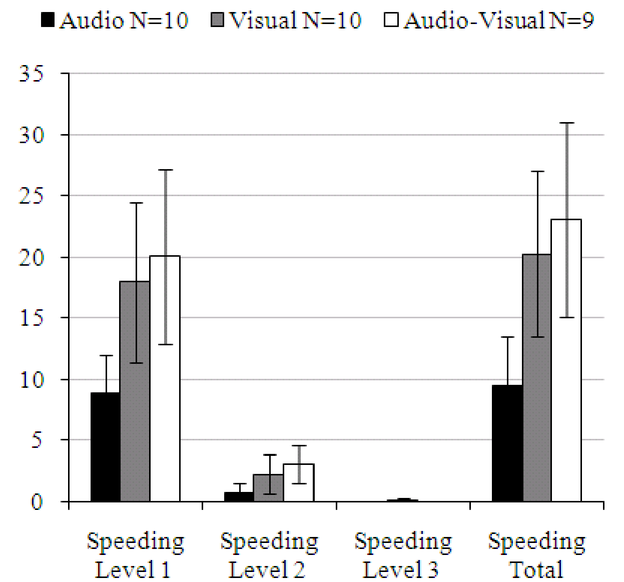
RESULTS

In this section we present the results from the field experiment. These results will be presented in the following order; primary driving task performance, secondary driving task performance and eye glance behaviour. The results were subjected to one-way independent-samples ANOVA tests and Tukey's HSD post-hoc tests. The results of these calculations are presented in Figure 1, 2 and 3.

Primary Driving Task Performance

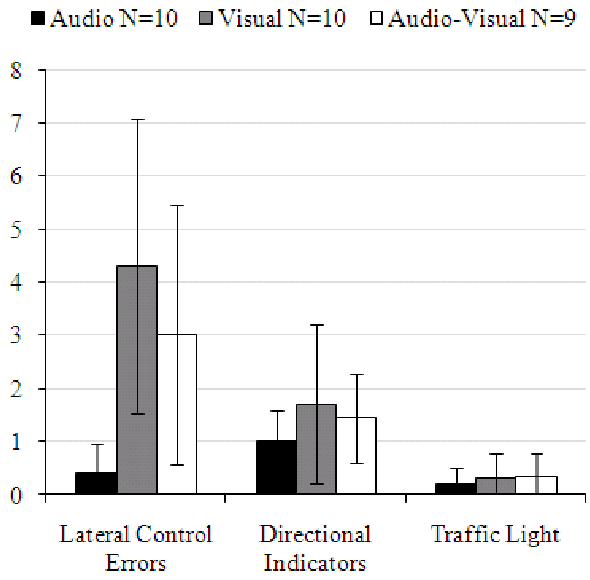
The variables for measuring primary driving task performance included longitudinal control errors, lateral control errors, directional indicator and traffic light violations. We identified a total of 648 violations concerning primary driving task performance across all three configurations (Audio = 111, Visual = 265, Audio-Visual = 272). Of the 648 primary driving task violations, 523 are classified as longitudinal control errors (see Figure 1). When assessing speeding violations we identified some major differences between the three configurations.

We identified a rather high number of speeding level 1 violations (exceeding the prescribed speed limit by three kilometres per hour) and our experiment showed that participants using the audio configuration on average had 8.8 (SD = 3.2) violations, participants using the visual configuration had 17.9 (SD = 6.5) and participants using audio-visual had 20.00 (SD = 7.11). An ANOVA test showed significant difference among the three configu-rations, $F(2,26) = 6.93$, $p = 0.004$. A Tukey's post-hoc test showed difference at the 5% level between visual and audio participants, with participants using the visual configuration having significantly more. When comparing speeding level 1 violations between the audio and audio-visual configurations a strong significant difference is revealed. Participants using the audio configuration committed fewer violations than participants using the audio-visual configuration, $p < 0.01$. A comparison between the visual and audio-visual configurations showed no significant difference.



Figur 1. Longitudinal control errors from the three configurations. The figure illustrates the average number of speeding violations.

When looking at speeding level 2 violations (exceeding the prescribed speed limit by 30 percent), we found that participants using the audio configuration on average had 0.7 (SD = 0.84) violations, while participants using visual had 2.2 (SD = 1.6) and participants assigned to the audio-visual configuration had 3 (SD = 1.56). An initial comparison of the results, reveals difference amongst the three configurations, where more than 50% of the identified speeding level 2 violations (Audio = 7, Visual = 22, Audio-Visual = 33), involved participants using the audio-visual configuration, $F(2,26) = 4.27$, $p = 0.025$. Not surprisingly, a post-hoc test confirmed that audio-visual participants had significantly more speeding level 2 violations when compared to audio participants, $p < 0.05$ – participants using audio-visual accounted for nearly five times as many speeding level 2 violations compared to audio participants,. We found no statistically significant differences between the audio and visual configurations or between the visual and audio-visual configurations. Looking at speeding level 3 violations (exceeding the prescribed speed limit by 60 percent) we found only one occurrence with one participant using the visual configuration – this was however not a significant difference.



Figur 2. Lateral control errors, directional indicators and traffic light violation incidents from the three configurations. The figure illustrates the average number of violations for each configuration.

If we look at all instances of longitudinal control errors (speeding violations, see Figure 1), we find that participants using the audio configuration had an average of 9.5 (SD = 3.9) incidents, participants using the visual configuration had 20.2 (SD = 6.8) and participants using audio-visual had 23 (SD = 8). An ANOVA test showed that there is a significant difference between the configurations, $F(2, 26) = 7.79$, $p = 0.002$. Furthermore we see that participants using the audio configuration had fewer speeding violations in total in comparison to participants using the visual and audio-visual configurations. This may not come as a surprise when the results from the experiment show that the visual (accounts for 202 violations) and audio-visual configurations (accounts for 226 violations) both have more than twice the speeding violations in comparison to the audio configuration (accounts for 95 violations). A Tukey's post-hoc test showed significant difference at the 5% level between audio and visual participants as well as strong significant difference at the 1% level between audio and audio-visual participants. When re-assessing the incidents related to longitudinal control we see that there is no significant differences between the audio-visual and visual configurations – both configurations constitute almost an equal number of violations.

We identified 77 lateral control errors (lane excursions, see Figure 2). When assessing the lateral control errors our experiment shows that participants using the audio configuration on average had 0.4 (SD = 0.56) incidents, participants using the visual configuration had 4.3 (SD = 2.76), while participants using audio-visual had 3 (SD = 2.44). When comparing the number of lateral control errors connected to participants within each of the configurations we see a major difference as participants in the visual and audio-visual configurations collectively constitute 95% of all incidents. Participants in the audio configuration only account for two incidents in total. An ANOVA test confirms this difference to be significant, $F(2,26) = 4.72$, $p = 0.018$. Using a post-hoc test we determined that this difference lies between the audio and visual configurations, where the participants using the visual configuration constitute a total of 43 lateral control errors, which is significantly more than the number of lateral control errors committed by audio participants, $p < 0.05$. Surprisingly enough there is no significant difference between the audio and audio-visual configurations despite the fact that the audio-visual configuration has over seven times as many incidents.

When comparing the results concerning violations related to traffic light (not adhering to the caution and stop signals assigned by traffic lights) and the directional indicators (failing to activate the directional indicators), the number of violations reveals no significant differences when subjected to an ANOVA test. In relation to the directional indicators, participants using the audio configuration had an average of 1 violation (SD = 0.6), participants using a visual configuration had 1.7 (SD = 1.5) and participants using the audio-visual configuration had 1.44 violations (SD = 0.84). When assessing traffic light violations, the results show that the number of incidents is almost equally distributed amongst the three configurations.

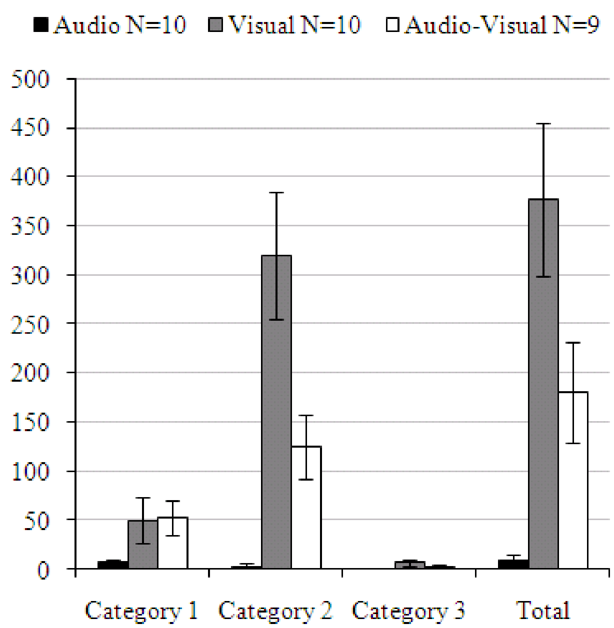
Secondary Driving Task Performance

When assessing secondary driving task performance, we used the following measurement variables; task completion time and navigational errors. Completion times for each of the four tasks were recorded and showed an average of 24.13 minutes (SD = 1.44) for audio users, 22.55 minutes (SD = 0.58) for visual and 23.05 minutes (SD = 1.06) for audio-visual users. We also identified a total of 34 navigational errors (Audio = 10, Visual = 14, Audio-Visual = 10). These results do not reveal any significant differences when compared across the three configurations.

Eye Glance Behaviour

We used three different variables for assessing eye glance behaviour; category 1 (< 0.5 seconds), category 2 (0.5 - 2 seconds) and category 3 (> 2 seconds). We identified a total of 5490 glances within all the field trials. When looking at our results we identified several major differences between the three configurations.

For the category 1 glances, our experiment showed that participants on average had 6.6 (SD = 3.6) in the audio configuration, 50.1 (SD = 23.12) in visual and 53 glances (SD = 17.11) in audio-visual (see Figure 3). An one-way ANOVA test confirms that there is an extremely significant differences amongst the three configurations, $F(2,26) = 13,19$, $p = 0.0001$. Subjecting these results to a Tukey's post-hoc test shows that participants using the audio configuration have significantly less category 1 glances than participants in the visual and audio-visual configurations, $p < 0.01$. This may not come as a surprise when the results show that audio participants only account for 6% of the recorded glances in the category 1 variable. The remaining occurrences in category 1 are almost equally distributed between the visual and audio-visual configurations; hence we found no significant differences.



Figur 3. Eye glance behaviour for the three configurations. The figure illustrates the average number of eye glances for each configuration.

When looking at category 2 glances (0.5 - 2 seconds), our experiment revealed a surprising total of 4358 glances (Audio = 28, Visual = 3201, Audio-Visual = 1129), with an average of 2.8 (SD = 3.6) for participants using the audio configuration, 320.1 (SD = 64.3) for participants using visual and 125.44 (SD = 32.72) for participants using audio-visual. Looking at the numbers for category 2 we can see that participants in both the visual and audio-visual configurations have an extremely high number of occurrences. Subjecting the results to an ANOVA test confirms that there are extremely significant differences amongst the configurations, $F(2,26) = 78,79$, $p = 0.0001$. A post-hoc test showed significant differences at the 1% level between the visual configuration and the two other configurations. This is also reflected in the results, where participants using the visual configuration accounted for 73% of all category 2 glances (three times as many as audio-visual participants), whereas audio participants accounted for less than 1%. The post-hoc test also revealed a significant difference between audio and audio-visual, where the audio configuration (comprise 28 incidents) has nearly none compared to audio-visual (comprise 1129 incidents), $p < 0.01$.

When assessing the category 3 variable (> 2 seconds), we found a total of 88 glances (Audio = 0, Visual = 67, Audio-Visual = 21), where audio-visual participants on average had 2.33 (SD = 2), participants using visual had 6.7 (SD = 3.7) and audio had 0. An ANOVA test showed that the difference amongst the configurations is significant, $F(2,26) = 12,61$, $p = 0.0001$. Participants using the visual configuration accounted for 76% of all category 3 glances. A post-hoc test revealed that participants using visual have significantly more category 3 glances than participants using the audio configuration, $p < 0.01$. A comparison of the visual and audio-visual configurations also reveals a significant difference, where participants using visual have more category 3 glances than audio-visual, $p < 0.05$. When looking at the category 3 results we see approximately the same glance ratio between the visual and the audio-visual configurations as in category 2 – participants using the visual configuration again accounted for three times as many glances as audio-visual participants.

Looking at the three glance categories collectively, we see that; 94 glances occurred in the audio configuration, 3769 in visual and 1627 in audio-visual. Participants using the audio configuration had an average of 9.4 (SD = 4.88), participants in visual had 376.9 (SD = 77.52) and audio-visual participants had 180.78 glances (SD = 51.31). When assessing the glance categories collectively we identified some major differences amongst all three configurations, which is also confirmed through an ANOVA test, $F(2,26) = 70.77$, $p = 0.0001$. The post-hoc test showed that the visual configuration, which accounted for 68.7% of all the glances, has a significantly higher number of glances compared to the audio configuration, which only accounted for 1.7% of the glances, $p < 0.01$. When comparing the audio and audio-visual configurations, we found that the audio configuration has a significantly lower number of glances than audio-visual, $p < 0.01$. When comparing the visual and audio-visual configurations a significant difference at the 1% level is revealed. Audio-visual participants accounted for 29.6% of all glances – nonetheless, participants in the visual configurations accounted for significantly more.

DISCUSSION

Previous research indicates that the use of GPS systems can result in decreased driving performance. We set out to evaluate three configurations of output modalities in order to shed light on how drivers are affected by such a highly output oriented device. Through our evaluation we also sought to identify potential design implications. We will initiate the discussion by focusing on eye glance behaviour, since eyes-off-the-road time is known to affect primary driving task performance [3, 9].

Not surprisingly, our results show a connection between visual output and the glance occurrences. Although we expected visual output to incite eye glances, we were surprised by how often participants diverted their visual attention from the road to look at the system. A similar field experiment, which investigates different combinations of output modalities for a conceptual navigational system, show a glance frequency of one glance every 8.5 seconds for visual interfaces [11], whereas our experiment revealed that participants using the visual configuration on average looked at the system every 4 seconds. Since driving performance decreases when visual demand increases [3], this is an interesting finding in relation to how distraction levels in manufactured GPS systems might differ from conceptual configurations used in studies. The high glance frequency could also be explained by the fact that the participants in our experiment belonged to a relatively young age group. Similar studies indicate that younger and older drivers differ in driving behaviour [12]. Another study by Green shows that younger drivers on average have a higher glance frequency compared to older drivers [11].

We did not see any specific patterns in relation to when participants looked at the system. The eye glance frequency did not appear to be influenced by a specific environment. Participants repeatedly looked at the system in both densely populated areas as well as long rural segments. We did however observe the consequences of diverting visual attention from the road to the system. Long glance durations combined with a high glance frequency would result in participants missing turns when attempting to relate the map provided by the GPS to their surroundings. Two of the participants became so engulfed by the system that the vehicle nearly came to a standstill while traversing a roundabout.

When comparing the total number of eye glances across the configurations we find that by introducing audio output to the system the number of glances is significantly reduced. The participants using the visual configuration looked at the system every 4 seconds on average, whereas participants using the audio-visual configuration had an average glance frequency of 7.5 seconds. However, if we consider that all participants in the experiment were equally successful in completing the navigational tasks, regardless of the configuration and the fact that participants using the audio configuration had a glance frequency of 2.5 minutes, a frequency of 7.5 is still extremely high. Having more than one output source available did not seem to be an advantage. Visual output only seemed to make participants take their eyes of the road, presumably because the participants found driving tedious and felt entertained by the visual output.

Combining visual and audio also resulted in a change in glance behaviour. During our data analysis we noted that all participants using the visual configuration repeatedly looked at the system while performing manoeuvres – for example, while making a turn or traversing a roundabout. This behavioural pattern only occurred in two field trials, which involved the audio-visual configuration. This could be explained by the difference in how visual and audio information is provided. Visual information is readily available allowing users to retrieve information whenever they deem it necessary. On the contrary, participants using the audio configuration can only receive information when the system deems it necessary. Since GPS systems have the role as the navigator within unfamiliar areas, this would suggest that participants depending solely on visual output constantly need to confirm their manoeuvres, the addition of audio output seems to alleviate this need.

While the frequency of glances differed significantly between the visual and audio-visual configurations, it becomes interesting when considering the results from the primary driving task performance variables. During our data analysis we noted several incidents, such as running red lights, missing turns and speeding caused by participants looking at the system. In spite of these consequences and what previous research [13] show, we found no significant differences between the driving performance of participants using the visual and audio-visual configurations. This comes as a surprise since decrease in primary driving task performance tends to be attributed to a lack of visual attention [1].

While the glance frequency did not seem to affect participants driving performance – the presence of visual output however did. Our results show that participants using the audio configuration performed better in relation to primary driving task performance than any of the other participants. These findings, which correspond with other studies [11, 17], indicate that an audio configuration would be most ideal in terms of road safety. Nonetheless, it is also important to consider that the differences in driving performance between the configurations could be explained by the difference in how visual and audio information is acquired. Cautious driving could be a result of participants relying on the GPS system to guide them and not being in control of when and where they receive instructions. This level of uncertainty and alertness could cause an increased cognitive workload of drivers. This matter is further emphasised by a behavioural pattern seen in all audio participants, where they drastically decrease speed when presented with auditory instructions. Studies on the effects of voice instructions reveal that a decrease in speed is one of the most significant indicators of increased cognitive workload [19].

A participant who used the visual configuration, intentionally diverted from the route assigned by the GPS system. This was motivated by the fact that the participant knew a better route. The participant defined a better route, as involving less turns and traffic lights, while permitting a higher speed level. The tendency to ignore GPS instructions while driving in familiar areas is also identified in a study by Leshed et al. The study describes how users would still utilize the system in order to feel in control by locating and orienting themselves on the map [16]. This pattern was also observed during this particular incident, where the participant repeatedly looked at the system in order to see if his chosen route was shorter than the one recommended by the GPS system.

When considering the measurement variables, our results show an inclination towards the audio configuration. To shed light on this matter we assessed the data collected through the post-task questionnaire in order to evaluate the participants' opinions of the audio configuration. When asked to assess the system instructions and output modality, the audio configuration was rated highest (rated from 'okay' to 'very satisfied') compared to visual and audio-visual (several of the ratings were 'dissatisfied' and 'very dissatisfied'). The outcome of the questionnaire indicates an overall satisfaction amongst the audio participants. The responses given in the interview contradict these findings. Over half the participants expressed that they would prefer the presence of visual output (where two only preferred visual output). Interestingly, half of the visual participants expressed the desire to have both visual and audio output (three preferred audio only), while half of the audio-visual participants would have preferred visual output only. The expressed opinions contradict each other, but a third of the participants stated that they would prefer to enable and disable the audio output in accordance to their own preferences. This seems to indicate that the audio-visual configuration would result in the highest user satisfaction, even though results indicated that it is less safe.

Design Implications

In this section we propose ideas of how future navigational systems could be designed to support drivers in navigating, while still taking the safety-critical aspects of the context into consideration.

Decreases in driving performance are primarily attributed to visual output. We believe an improvement in relation to road safety could be attained by restricting or omitting the visual output when in motion.

Several participants that used the audio configuration expressed a need for more output during long straight road segments. The participants also expected more guidance before they were to traverse a complex intersection. Allowing the user to retrieve information whenever they deem it necessary, could pose as a potential solution, as it also maintains an auditory interface.

During the field trials, which involved the audio configuration, we observed that participants looked at the system during travel, even though no visual output was provided. Since many GPS systems are bought as an independent component, they have the disadvantage of giving the user a visual focal point that may attract their attention when presented with instructions. We believe that by integrating GPS systems into vehicles and utilizing the car stereo for audio output, could eliminate the visual focal point. This concept is similar to many hands-free phone systems.

A major point of criticism concerning the audio output from the GPS system was its use of keywords and metric distances in the navigational instructions. Over half of the participants with audio output experienced problems relating such information to their surroundings and requested use of more landmarks and descriptive details in the instructions. Studies have shown that direction is the most requested information when navigating in addition to landmarks, road numbers and street names, whereas distance is a less desired information [5]. Another study [16] also recommends the use of landmarks to support drivers. We believe that this additional descriptive information would not only support drivers, but also alleviate confusion caused by technical limitations in GPS systems. Over half of the participants in our experiment experienced problems with inconsistency in the GPS system's distance estimates – for example, delayed updates of visual maps or instructions due to unstable satellite signals or loss hereof. Areas, which had several navigational options, caused confusion amongst the drivers, which consequently led to navigational errors. Utilizing landmarks and additional descriptive information would further enable drivers to relate visual maps and instructions to their surroundings.

Limitations

Although we strived to approximate a natural setting, we cannot eliminate the possibility that participant behaviour was affected by the fact that they were being observed. During the concluding interview, two participants expressed that they chose to follow the provided route despite disagreements with the given instructions, as they believed that a linear approach was necessary in order to complete the tasks. This gives rise to speculations of how often the participants decided to follow the navigational instructions provided by the GPS system, even if the participants disagreed with the provided information and if this behaviour could have affected our results. We acknowledge that there is an imbalance between the visual and audio configurations in relation to the way the navigational information is provided. Participants using configurations with visual output had additional information, such as street names, estimated arrival time and distance, which was not available through auditory output. Only one participant commented on this, yet it is unknown if this imbalance could have affected the outcome of the experiment.

CONCLUSION

Research on in-vehicle systems indicates a need for further elaboration on how output modalities affect driving performance. In this study we compared three output configurations of a GPS system in order to shed light on how drivers are affected by such a highly oriented device. Our results indicate that visual output not only causes a substantial amount of eye glances, but also leads to a considerable decrease in driving performance. While the introduction of audio output in combination with visual output reduced the frequency of glances, we were surprised to discover that this did not have any effect on the driving performance. This could indicate that the presence of audio output increases the cognitive workload, nonetheless the audio configuration is still ideal when considering eye glance behaviour and glance tendencies. Although the audio-only configuration proved to be the most favourable in relation to driving performance, the user satisfaction inquiries show a preference for having both output modalities available.

A direction to be pursued is to design an audio output based navigation system, which accedes to user preferences, as our results already indicate that audio output is an adequate output modality in terms of road safety. Moreover, further study is needed to fully understand the behavioural patterns, which emerge when using GPS systems.

REFERENCES

1. Bach, K. M., Jæger, M. G., Skov, M. B., and Thomassen, N. G.. 2007. A Classification of In-Vehicle Systems Research: Understanding, Measuring and Evaluating Attention, HCI Lab Technical Report.
2. Bach, K. M., Jæger, M., Skov, M. B., G., Thomassen, N. G. 2008. You Can Touch, but You Can't Look: Interacting with In-Vehicle Systems. Proceeding of the 26th annual SIGCHI conference on Human factors in computing systems CHI 08, ACM.
3. Brooks, C. & Rakotonirainy, A. 2007. In-vehicle technologies, Advanced Driver Assistance Systems and driver distraction: Research challenges. Distracted driving. Australasian College of Road Safety.
4. Broy, M. 2006. Challenges in Automotive Software Engineering in ICSE, ACM Press.
5. Burnett, G. (2000) Turn right at the traffic lights - The requirement for landmarks in vehicle navigation systems, The Journal of Navigation.
6. Cassavoy, L. TOMTOM GO 930 Review, PC World, November 2008.
7. Devonshire, J., Flannagan, C., and Sayer, J. 2005. The Effects of Secondary Tasks on Naturalistic Driving Performance, The University of Michigan Transportation Research Institute, UMTRI-2005-29.
8. Geiger, M., Zobl, M., Bengler, K., Lang, M., Intermodal differences in distraction effects while controlling automotive user interfaces. Proceedings Usability Evaluation and Interface Design, HCI 2001.
9. Green, P. 1996. Customer Needs, New Technology, Human Factors, and Driver Science Research for Future Automobiles in Journal of the Society of Mechanical Engineers, University of Michigan Transportation Research Institute.
10. Green, P. 1999. Visual and Task Demands of Driver Information Systems, The University of Michigan Transportation Research Institute.
11. Green, P., Hoekstra, E., Williams, M., Wen, C., and George, K. 1993. Examination of a videotape-based method to evaluate the usability of route guidance and traffic information systems, UMTRI-93-31.
12. Horberry, T., Anderson, J., Regan, M. A., Triggs, T. J., and Brown, J. 2006. Driver distraction: The effects of concurrent in-vehicle tasks, road environment complexity and age on driving performance, Accident Analysis and Prevention.
13. Jovanis, P., Srinivasan, R., Effect of in-vehicle route guidance systems on driver workload and choice of vehicle speed: Findings From a Driving Simulator Experiment in Ergonomics and safety of intelligent driver interfaces, 1997.
14. Keith, O., and Burnett, G.. 2008. Learning-oriented Vehicle Navigation Systems: A Preliminary Investigation in a Driving Simulator, MobileHCI 2008.
15. Lee, D., Caven, B., Haake, S., Brown, T. 2001. Speech-based Interaction with In-vehicle Computers: The Effect of Speech-based E-mail on Drivers' Attention to the Roadway, University of Iowa.
16. Leshed, G., Velden, T., Rieger, O., Kot, B., Sengers, P. 2008. In-Car GPS Navigation: Engagement with and Disengagement from the Environment, CHI 2008.
17. McCrickard, S., Moldenhauer, D., Effect of Information Modality on Geographic Cognition in Car Navigation Systems. Center for HCI and Department of Computer Science, Virginia Tech.
18. Nixon, C., Anderson, T., Morris, L., McCavitt, A., McKinley, R., Yeager, D., and McDaniel, M. 1998. Female voice communications in high level aircraft cockpit noises part II: vocoder and automatic speech recognition systems. Aviation, Space and Environmental Medicine.
19. Uchiyama, Y., Kojima, S., Hongo, T., Terashima, R. and Wakita, T. 2002. Voice Information System Adapted to Driver's Mental Workload. Proceedings of the Human Factors and Ergonomics Society.
20. Zellweger, P., Bouvin, N., Jehøj, H., and Mackinlay, J.D. 2001 Fluid Annotations in an Open World. Proc., ACM Press.

Appendix D, Field Experiment