



HAL
open science

Don't Look at Me, I'm Talking to You: Investigating Input and Output Modalities for In-Vehicle Systems

Lars Holm Christiansen, Nikolaj Yde Frederiksen, Brit Susan Jensen, Alex Ranch, Mikael B. Skov, Nissanthen Thiruravichandran

► To cite this version:

Lars Holm Christiansen, Nikolaj Yde Frederiksen, Brit Susan Jensen, Alex Ranch, Mikael B. Skov, et al.. Don't Look at Me, I'm Talking to You: Investigating Input and Output Modalities for In-Vehicle Systems. 13th International Conference on Human-Computer Interaction (INTERACT), Sep 2011, Lisbon, Portugal. pp.675-691, 10.1007/978-3-642-23771-3_49 . hal-01590843

HAL Id: hal-01590843

<https://inria.hal.science/hal-01590843v1>

Submitted on 20 Sep 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

Don't Look at Me, I'm Talking to You: Investigating Input and Output Modalities for In-Vehicle Systems

Lars Holm Christiansen, Nikolaj Yde Frederiksen, Brit Susan Jensen,
Alex Ranch, Mikael B. Skov, Nissanthen Thiruravichandran

HCI Lab, Department of Computer Science, Aalborg University
Selma Lagerlöfs Vej 300, 9220 Aalborg East, Denmark
{lhc_dk, alex_ranch}@hotmail.com, nyf@mail.dk,
{britjensen, nissanthen}@gmail.com, dubois@cs.aau.dk

Abstract. With a growing number of in-vehicle systems integrated in contemporary cars, the risk of driver distraction and lack of attention on the primary task of driving is increasing. One major research area concerns eyes-off-the-road and mind-off-the-road that are manifested in different ways for input and output techniques. In this paper, we investigate in-vehicle systems input and output techniques to compare their effects on driving behavior and attention. We compare four techniques touch and gesture (input) and visual and audio (output) in a driving simulator. Our results showed that the separation of input and output is non-trivial. Gesture input resulted in significantly fewer eye glances compared to touch input, but also resulted in poorer primary driving task performance. Further, using audio as output resulted in significantly fewer eye glances, but on the other hand also longer task completion times and inferior primary driving task performance compared to visual output.

Keywords. In-vehicle systems, touch interaction, gesture interaction, eye glances, driving performance

1 Introduction

Driver attention is critical while drivers handle vehicles and interacts with emerging technologies in contemporary cars. Attention is the human's ability to concentrate on certain objects or situations and allocate processing resources accordingly [8]. But as technologies like GPS navigation systems or in-car media players are increasingly being used in cars, attention becomes an important aspect to consider when designing in-vehicle systems. With the driver's attention being divided between in-vehicle systems and the driving task, risks of accidents increase [4, 9]. A fundamental problem is that the driver has to remove attention from the primary task of driving to performing secondary tasks such as interacting with the car stereo. As the amount and complexity of in-vehicle systems increase, so does the demands on driver attention. This introduces challenges when trying to achieve safe and efficient interaction in order to minimize the amount of time the driver has to remove attention from the road in particular the driver's visual attention [2, 3].

Technological progress and price reductions have pushed a growing use of touch-based interfaces to control various kinds of in-vehicle systems. The flexibility in its application capabilities, low price, and utilization of a more natural way of interaction, makes it an obvious choice for in-vehicle systems with its increasing presence in new cars and aftermarket GPS-units. But the inherent characteristics of touch screens imply high requirements for visual attention of the driver due to the lack of immediate tactile feedback or dynamics of screen layout [3]. This leads to withdrawal of attention from the driver [25].

Brown [5] classifies two types of withdrawal of attention namely general and selective. The general withdrawal of attention refers to insufficient visual perception during the driving situation. This is also known as eyes-off-the-road [11]. Interacting with in-vehicle systems often requires selective withdrawal of attention as drivers read displays or push buttons. Selective withdrawal of attention is a more subtle distraction type as it deals with mental processing, e.g. memory processes or decision selection. It is also known as mind-off-the-road [12] and takes place e.g. while talking to other passengers or while talking on the phone. Interacting with in-vehicle systems can lead to mind-off-the-road while interacting with e.g. a speech system or listening to audio instruction [2]. Thus, we can distinguish between input and output techniques that require high or low visual attention and perhaps lead to eyes-off-the-road or mind-off-the-road.

Inspired by previous research on touch-screen interaction technologies [3, 20], we compare different input and output techniques for in-vehicle systems to investigate and measure their effects on the driving activity and driver attention. The paper is structured as follows; initially we present previous research on in-vehicle systems, and secondly we introduce the interaction techniques. Then we describe the experiment and the results are presented and finally, we discuss and conclude the results.

2 Related Work

Driver attention and distraction are fundamental concepts when doing research or development within vehicle safety or in-vehicle systems design [2]. Attention can be defined as the ability to concentrate and selectively focus or shift focus between selected stimuli [8, 18]. Within cars and other vehicles, driver attention is primarily focused on monitoring the environment and executing maneuvers also called the primary driving task [2, 6, 13, 16]. Disruption of attention is defined as distraction and Green describes distraction as anything that grabs and retains the attention of the driver, shifting focus away from the primary driving task [2, 4, 13].

Within in-vehicle attention and distraction research, we observe a significant focus on the dynamics between the primary driving task and secondary driving tasks, e.g. operating various in-vehicle systems. This is significant since research identifies the use of in-vehicle systems as a cause of traffic accidents [4, 13]. Green [13] stresses that most drivers will go to great lengths to complete a given secondary task and rarely abandon a task upon initiation. With a critical primary task, this seemingly irrational behavior and distribution of attention between the primary and secondary

task can endanger the safety of the driver and the surroundings. Lansdown et al. [16] acknowledge this unsettling tendency concerning in-vehicle systems in a study focusing on driver distraction imposed by in-vehicle secondary systems.

A tendency within in-vehicle interaction research involves attempts to identify an interaction technique that surpasses the capabilities of the traditional tactile interface. In a comparative study, Geiger et al. [9] set out to evaluate the use of dynamic hand movements (gestures) in order to operate a secondary in-car system and compare it to a traditional haptic (tactile) interface. The parameters used for comparison, were errors related to driving performance, tactile/gesture recognition performance and the amount of time drivers didn't have their hands on the steering wheel. The experiment showed that use of the tactile interface resulted in high task completion times and the system lacked in recognition performance when compared to the gesture interface. The gesture interface allowed users to perform the primary task appropriately, while the users also found the gesture interface more pleasant and less distracting. Alpern & Minardo [1] support these findings in a study where they evaluated gestures through an iterative development of an interface for performing secondary tasks. In the final iteration of their experiment, they noted that users made fewer errors compared to a traditional tactile radio interface. Findings from both studies indicate that gestures could be a viable alternative for secondary in-car systems.

Bach et al. investigated how perceptual and task-specific resources are allocated while operating audio systems in a vehicle [3]. Three system configurations – a conventional tactile car stereo, a touch interface and an interface that recognizes gestures as input – were evaluated in two complementary experiments. Bach et al. identified an overall preference for the gesture-based configuration as it enabled the drivers to reserve their visual attention for controlling the vehicle. The conventional car stereo on the other hand lacked an intuitive interface; consequently the system required high perceptual and task-specific resources to be operated affecting the primary task performance of the subjects. The touch interface introduced a reduction in overall task completion time and interaction errors when compared to both the conventional tactile and gesture interfaces. While the potentials of gestures as an input method for in-vehicle systems seems promising, little attention has been given to the possible influence of output methods. In order to address this it would be necessary to distinguish between input and output to clarify how combinations of different output and input methods might affect the interaction and primary task performance.

The need to separate output from input in relation to in-vehicle systems is acknowledged by Bach et al. [3] as a limitation in their study and the need for further research on this topic is recognized. Their initial research focus was on system input as opposed to output, which meant the included output mechanisms differed for each of their configurations. The variation in output could have affected the findings – the results do not show which kind of output mechanism that is suitable for in-vehicle systems. This suggests an additional study on output methods in order to investigate how they influence primary task performance and secondary task performance in the vehicle domain.

The aim of our study is to compare different configurations of in-vehicle systems with an equal emphasis on both input and output mechanisms. We aim to confine system variables regarding input and 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 the interaction possibilities of touch screens, which support both touch and gesture interaction and the polarity in the two different sensory channels of output.

3 In-Vehicle System: Input and Output

We distinguish between input and output in this experiment while using a touch-screen. We integrate two different kinds of input and two different kinds of output enabling four different in-vehicle configurations; touch input with visual output, touch input with audio output, gesture input with visual output and gesture input with audio output. These configurations will hereafter be referred to as <input>/<output>, e.g. touch/visual.

In order to evaluate these configurations with regards to their effect on attention, we chose a well-known in-vehicle system as case system namely a music player or car stereo. This decision was based on previous studies, e.g. [1, 3, 13], and the music player served as a somewhat simple yet sufficient platform for our experiment. The system was designed for an 8" 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 transparent plastic shield to discourage deliberate input and prevent accidental input in this area.

3.1 Input: Touch-Based and Gesture-Based

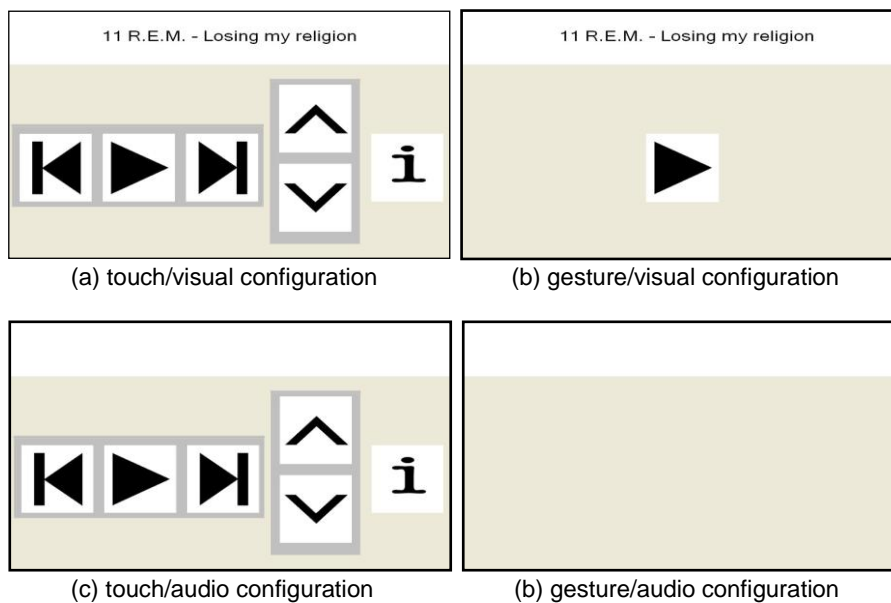
We integrated two different input methods - conventional touch-based input with graphical buttons and gesture-based input using the touch-screen as a drawing canvas.

The layout of the two touch configurations was inspired by Bach et al. [3] and our goal was to keep it as simple as possible, while still providing the necessary basic functionality for controlling the music player. The implemented icons on the interface resemble common music player icons to minimize interpretation. The buttons were furthermore grouped according to 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 keep the design consistent. The size and spacing of the buttons was inspired by previous research on touch screen layout, e.g. [7, 22, 23]. Input is only possible by pressing the buttons according to the click-on-release principle. This means that the buttons are activated only when the finger has left the button, which also means that nothing happens when a button is held.

The gesture-based systems have no buttons. Instead, the systems are controlled by gestures drawn directly on the screen using a finger. The gestures used are inspired by Pirhonen et al. [20] and Bach et al. [3] and allow for the same functionality as the touch buttons. The only gesture that is different 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" often used as an icon for "information". The player is controlled through gestures anywhere on the input area of the canvas. The gesture system features the same functionality as the the

functionality on the touch-based interface, e.g. play, pause, skip forward or backward, volume up or down.

Figure 1. The graphical user interface for the four configurations (a) touch/visual, (b) gesture/visual, (c) touch/audio, and (4) gesture/audio. The top part of the screen (white) 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.



3.2 Output: Visual-Based and Audio-Based

We integrated two different output methods namely visual using icons and text and audio using ear-cons and voice. Visual and audio output is not used simultaneously at any point. We further distinguish between two kinds of output feedback on input and information about the state of the system.

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 (figure 1.a), this is done by changing the appearance of buttons to indicate they have been pressed. Furthermore, when the volume is all the way down, pressing the “Volume down” button will change its appearance to reflect a disabled state. The same principle applies to the “Volume up” 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 about one second (as shown on figure 1.b). Audio feedback is implemented using ear-cons. When the user either pushes a

button or performs a gesture (for the touch/audio and gesture/audio configurations), 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 results in a “dong” sound.

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 equivalent audio output is implemented using playback of voice recordings containing the same information. Either pushing the “Song info” button or performing the “Song info” gesture plays these recordings.

4 Experiment

The purpose of the experiment was to compare the four different configurations of the system (visual, audio) x (touch, gesture) and consequently the different ways for in-vehicle interaction.

4.1 Experimental Design

We adapted a between-subject design with 32 participants with four groups of eight subjects corresponding to the four configurations. Each group consisted of four male and four female test subjects and was assigned to one of the four configurations of our music player.

4.2 Subjects

32 people (16 males and 16 females) in the age 21 to 56 years old ($M=28.2$, $SD=9.2$) participated in our experiment. They all stated that they were in good health condition. All of the test subjects carried valid driver’s licenses and had done so for between 0.5 and 29 years ($M=9.4$, $SD=8.7$). Their driving experience was quite varied and ranged between 100 and 30.000 km/year ($M=6114.7$, $SD=7989.9$). Two test subjects stated that they had previous experience with the driving simulator used in our experiment. Subjects were balanced between the conditions according to gender, age, and driving experience.

4.3 Setting

We created a medium-fidelity driving simulator (similar to [3, 15, 17]) at our HCI laboratory at Aalborg University. We adapted two car seats, steering wheels (with force feedback) and a brake and accelerator pedal using the application Test Drive Unlimited (TDU) running on a desktop computer (see figure 2). The driving application TDU was chosen as it provides a driving context that shares similar characteristics as real-traffic driving, e.g. driving in towns and rural areas introducing

unexpected potential hazards as the application integrates computer-controlled drivers on the road (similar to [19]). The setup also included two sets of speakers with one set of 4.1 surround sound speakers playing the sound of the game, and another set of 2.1 stereo speakers for music playback. The game was projected onto the wall in front of the subjects. The speedometer and tachometer of the car was visible to the subjects during the test as part of the projected image. The test subjects occupied the driver's seat while the test leader sat in the passenger seat during the test (figure 2.3).

Figure 2. Excerpt from the video recordings illustrating the driving simulator and the four different cameras. (1) Capturing eye glance for driver attention analysis, (2) input and output screen (here for gesture and audio), (3) the driver and experiment facilitator, and (4) (left) the driving simulator.



4.4 Tasks

The subjects were asked to solve 32 tasks during the test. Half of the tasks focused on system input, while the other half on system output. Furthermore, we attempted to create the tasks in such a manner that they did not favor any of the four configurations. The tasks were chosen to reflect realistic interactions with an in-car music player, e.g. changing tracks on a CD. Instructions for each task were kept short and clear in order to minimize interruption. The tasks varied in complexity ranging from simple ones like “Stop the music” to more demanding ones like “Find and play the song by Coldplay called Viva la Vida”. The tasks were all read aloud by the test leader.

4.5 Procedure

All sessions followed the same basic procedure. First, we collected demographical data for the test subjects. Then, the subjects were asked to take a seat in the simulator and make sure that the driving position was comfortable. The test leader then briefed the participants by reading a text aloud, which told them what they were about to do. They were also shown how to operate the music player in the particular configuration they were to use during the experiment. After each instruction was demonstrated, the subjects were asked to repeat it, in order to ensure they had understood how to operate the system. The subjects were instructed to drive the car between 40 and 60 km/h, except when performing maneuvers like turning and braking, to stay in the right lane, and otherwise observe normal traffic regulations and drive as they would in a real car. The subjects were then given a chance to familiarize themselves with the driving simulator and the steering wheel and pedals as they were allowed to try the game prior to the test itself.

After the practice run the test leader reset the simulator and the actual test session began. The driving itself was divided into two parts. In the first part the test leader would instruct the subjects where to turn, making sure they all followed the same predetermined route. In the second part, the subjects were told to drive freely in the environment. The length of each part was determined by the tasks, which they were asked to solve while driving. The tasks were divided evenly between the two parts, with 16 tasks to be solved in each. The subjects were instructed to start solving the tasks only when they felt ready to do so. The test sessions were recorded on four different video cameras for later analysis (as illustrated in figure 2).

4.6 Data Analysis

Inspired by [2, 3, 16, 26] we integrated several dependent measures to assess driver attention and driving performance while participating in each of the configurations. We chose to apply the following:

1. Primary driving task performance
2. Secondary driving task performance
3. Eye glance behavior

1) Primary driving task performance was measured as the number of errors in lateral and longitudinal control (i.e. [1, 3, 16]). A lateral control error was defined as a lane excursion where the subject failed to stay within the two lines denoting the right hand side lane of the road. Longitudinal control errors were defined as failure to maintain a speed within the instructed range of 40-60 km/h. A longitudinal error was noted each time the subject went above or below the speed range. Thus, staying at a wrong speed for a period of time only counted as one driving error. We identified these driving errors through video examination (elaborated below).

2) Secondary driving task performance we defined as interaction errors and task completion time (commonly applied in in-vehicle research [1, 16]). Interaction errors were defined as attempts to interact with the system that either had no effect or didn't

have the effect towards completion of the task that the subjects expected. In order to identify these errors, one of the cameras recorded an up-close view of the interaction with the screen. Task completion time was measured from the time the subjects 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.

3) Eye glance analysis is a highly used metric for analyzing driver attention within in-vehicle research [2, 3, 10]. We divided into three categories according to duration as category 1 was an eye glance below 0.5 seconds, category 2 was a eye glance between 0.5 and 2.0 seconds, and category 3 was an eye glance above 2.0 seconds. The nature of particular the eye glance analysis meant that it was necessary to view the videos frame by frame. In determining the length of an eye glance for instance, we knew that the each second of video contained 25 frames and a glance of 0.5 seconds or less corresponded to 12.5 frames (in practice 13 frames).

We analyzed two randomly picked sessions in order to achieve agreement in the interpretations of the data. This gave us the opportunity to discuss the various types of incidents in the data and we compiled a list of guidelines for the individual analyses. Each of the 32 sessions was analyzed by three of the authors of this paper. Each reviewer analyzed the video individually while logging and categorizing instances of all the abovementioned incidents. The resulting three logs were then compared and compiled into one final list containing all the incidents for that session. This was done by way of majority vote; if for instance only one reviewer had recorded a specific incident, which neither of the two other reviewers had recorded, the incident would not make it to the final list, and so forth. The same principle applied to categorization of eye glances. In situations where no majority vote could be secured, the video recording was reviewed again in order to reach the final decision.

5 Results

The results of the data analysis are presented in three sections namely primary driving task performance (including lateral and longitudinal control), secondary driving task performance (interaction errors and time) and eye glance behavior.

In each section, we first compare the results for the two input methods (N=16), then the two output methods (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 ANOVA tests as well as Tukey's HSD post hoc tests where applicable. The data is organized into two tables; one for N=16 (table 1) and one for N=8 (table 2). Any statistically significant differences are highlighted.

5.1 Primary Driving Task Performance

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

Table 1. Primary driving task performance across the input and output configurations (N=16). A plus denotes a significant difference at a 5% significance level.

	Input		Output	
	Touch (N=16)	Gesture (N=16)	Visual (N=16)	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)

When comparing the primary driving task performance across the two input methods, we see no significant difference between any of the metrics, although gesture input generally has a higher number of errors across all the metrics. Looking at the results of the primary task performance for the output methods, however, does reveal a significant difference in the number of speed increases, with visual having significantly fewer than audio, $t=2,04$, $p<0.05$. However, there are no significant differences in the number of total 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.

Table 2. Primary driving task performance across the four configurations (N=8). A plus denotes a significant difference at a 5% significance level.

	Touch Visual (N=8)	Touch Audio (N=8)	Gesture Visual (N=8)	Gesture Audio (N=8)
	Lane excursions	7.63 (4.87)	6.75 (5.01)	9.63 (8.28)
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)

Considering the primary driving task performance results and compare the four configurations (see table 2), 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 vs. the gesture/audio configuration ($p<0.05$). The remaining measurements of primary driving task performance show no significant differences. But the results do show that the two audio configurations have the highest number of total speed deviations.

5.2 Secondary Driving Task Performance

For secondary driving task performance we measured the total task completion time and identified a total of 1018 interaction errors. Comparing just input methods, 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, $t=2.04$, $p<0.19$.

Table 3. Secondary driving task performance across the input and output configurations (N=16). A plus denotes a significant difference at a 5% significance level.

	Input		Output	
	Touch (N=16)	Gesture (N=16)	Visual (N=16)	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) +

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

Table 4. Secondary driving task performance across the four configurations (N=8). A plus denotes a significant difference at a 5% significance level.

	Touch Visual (N=8)	Touch Audio (N=8)	Gesture Visual (N=8)	Gesture Audio (N=8)
	Interaction errors	42.38 (19.72)	16.38 (7.46)	39.00 (37.72)
Task completion time	249.88 (24.28) -	292.13 (81.62)	264.00 (95.42)	353.63 (75.66) +

Secondary driving task performance results reveal no significant differences in the number of interaction errors distributed among the four configurations, even though the average number of interaction errors for the touch/audio configuration is less than half that of the 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 for the touch/visual and gesture/audio configurations ($p<0.05$).

5.3 Eye Glance Behavior

We identified a total of 2371 glances divided into 560 glances below 0.5 seconds, 1729 between 0.5 and 2.0 seconds and 52 above 2.0 seconds. Of the total glances, around 60% occurred with touch input, which amounts to a significant difference 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 glances between 0.5 and 2.0 seconds, with gesture input having substantially fewer, $t=2.04$, $p<0.01$. But in the two remaining categories touch has the fewest, although the difference is only marginal.

Table 5. Eye glance behavior across the input and output configurations (N=16). A plus denotes a significant difference at a 5% significance level.

	Input		Output	
	Touch (N=16)	Gesture (N=16)	Visual (N=16)	Audio (N=16)
< 0.5 s.	16.44 (13.85)	20.44 (12.09)	15.94 (11.85)	20.94 (13.88)
0.5 – 2.0 s.	71.88 (19.35) +	36.19 (36.66) -	76.06 (24.34) +	32.00 (13.88) -
> 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) -

The number of glances for visual output account for 1523 (64%) of the total number of glances across output types, which amounts to an extreme significant difference, $t=2.04$, $p<0.001$. There is also an extreme significant difference in the number of glances between 0.5 seconds and 2.0 seconds with audio being significantly lower than visual, $t=2.04$, $p<0.001$. Finally, there also exists a strong significant difference in the number of glances above 2.0 seconds, with visual again having more (with 51 glances vs. just 1 glance), $t=2.04$, $p<0.01$. On the other hand, audio output has more glances below 0.5 seconds than visual output, albeit only marginally.

Across the four configurations, the touch/visual configuration accounts for around 32% of the total amount of glances, 27% for touch/audio, 31% for gesture/visual and just 8% for gesture/audio. A one-way ANOVA showed this difference to be extreme significant, $F(3, 28)=13.59$, $p<0.001$. Looking at these percentages, it is perhaps 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 glances below 0.5 seconds compared to e.g. touch/audio, this does not represent a significant difference, but a one-way repeated-measures ANOVA indicates that it is approaching significance, $F(3, 28)=2.65$, $p<0.07$. For glances between 0.5 and 2.0 seconds, however, an extreme significant difference exists, $F(3, 28)=30.22$, $p<0.001$. The results of the 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 glances between 0.5 and 2.0 seconds for touch/visual and touch/audio, $p<0.05$. In the last category, glances above 2.0 seconds, our results show an extreme 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 $p<0.01$ compared to touch/audio (0 glances) and gesture/audio (1 glance), and $p<0.05$ compared to touch/visual.

Table 6. Eye glance behavior across the four configurations (N=8). A plus denotes a significant difference at a 5% significance level.

	Touch Visual (N=8)	Touch Audio (N=8)	Gesture Visual (N=8)	Gesture Audio (N=8)
< 0.5 s.	8.88 (4.19)	24.00 (16.20)	23.00 (13.02)	17.88 (11.34)
0.5 – 2.0 s.	86.5 (12.40) + (+)	57.25 (12.62) + (-)	65.63 (29.44) +	6.75 (5.70) -
> 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) -

6 Discussion

The overall problem we set out to investigate was how to design in-vehicle systems that require as little visual attention from the driver as possible in order to avoid a decrease in driving performance, as current conventional techniques tend to do [16]. In the following we discuss and reflect on our results.

We were inspired by Bach et al. [3] who raise a question on the effects of separating input and output. This is what we have pursued 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 eye glances when comparing across output technique. This seems to imply that when conducting experiments with in-vehicle systems it is important to isolate and focus on both the input and output methods of the system.

6.1 Input

We initially assumed that touch-based input would require more eye glances than gesture input, as drivers would need to visually obtain the position of the buttons before commencing interaction. This was supported by our findings where we found a strong significant difference in glances between 0.5 and 2.0 seconds, and a significant difference in the total number of glances, which is in line with the work by [1, 20]. 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 glances between 0.5 and 2.0 seconds isolated, where touch input accounts for almost twice as many glances (98%) as gesture input. This is in line with Alpern & Minardo's findings which show that gesture interfaces, although not attention free, help drivers solve their task while allowing them to keep their eyes on the road [1].

The difference in eye glance behavior can perhaps 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 quick succession, users might have a tendency to use more glances in order to ensure/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 screen before being able to issue one or more commands without looking, as opposed to finding the correct button on the screen. This could be part of the explanation for the difference in the number of glances.

Before conducting the experiment we further assumed that gesture input would have relatively more glances below 0.5 seconds compared to touch, the rationale being that the aforementioned visual confirmation of the position of the screen 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 the findings of [3] our results also show touch as the fastest of the two input forms, although not significantly.

6.2 Output

We found some differences between the audio and visual output configurations when comparing measurements of primary driving task performance. Only in the number of speed increases is this difference significant in favor of visual output. However, the total number of speed deviations is not significantly different, so what these results indicate, if anything, is unclear since the number of speed decreases is almost identical, and the total amount of speed deviations imply no significant difference.

When comparing task completion time for the two output techniques of our system, there is a significant difference between the two, with visual output being faster. We believe this is due to the nature of audio output. When solving tasks requiring audio output, the user first has to hear the audio message, which can be of arbitrary length, and then process the information they are presented with before being able to solve the task. With visual output the user only has to read the information before being able to answer, which presumably takes less time. Or perhaps the user has already seen the information while performing another task, which further decreases the time required solving 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. Obviously, 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 clearly 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. As a result, 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.

In terms of road safety it can be argued that the increase in task completion time is a favorable tradeoff if it comes with fewer eye glances, which in turn leads to more

attention on the road. Our results do not however show a link between the number of glances and primary driving task performance, similar to the findings in [3]. However, other studies state that a relationship between eye glance behavior and driving performance do exist [10, 21]. In line with Gellatly [10], it is not difficult to imagine that more visual attention on the road is preferable, since the driver's primary method of assessing danger signs in traffic arguably is through the eyes. However, further studies are required in order to determine whether this is really the case. This is also indicated in a study on the effects of hands-free mobile phone conversations on driving performance [24]. Strayer and Drews state that even if drivers conducting a hands-free mobile phone conversation direct their gaze at the road, they often fail to notice objects in the driving environment, since their attention is occupied with conducting the mobile phone conversation. However, their findings relate to mobile phone conversations, which they claim might 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 marginally higher for audio output than for visual output. This could indicate that listening to audio output while driving 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, i.e. listening to voice messages from a hands-free mobile phone, results in a deterioration of driving performance [14]. Cognitive workload is also discussed in [3] in relation to their gesture/audio system, but their setup does not allow them to see an explicit connection to the output method, which leads them to attribute it to memory load, e.g. the driver having to remember the 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 show that the subjects receiving audio output spent significantly more time on completing the tasks. Hence, while audio output might result in fewer glances, the driver is occupied with the task for a longer time, if only mentally.

6.3 Limitations

Some of our participants found the limited level of realism in the simulator problematic. They pointed to the absence of tire noise, lack of opportunity to orientate themselves through the side and rear windows and sensation of movement, as some of the factors that they felt affected the realism and their driving performance. This was in part because these factors provide the drivers with a sensation of movement, which helps them estimate speed, without having to look at the road ahead. This could imply that particularly longitudinal control performance suffers from simulated driving [3].

Our choice of case system represents a possible source of inaccuracy. The nature of the music player means that it will always give a form of audio feedback, regardless of which output methods we choose. For instance, pushing the “Play” button will

cause music to be played; turning up the volume will cause the music to become louder, etc. This means that test subjects given visual output would not necessarily need to look at the screen to receive feedback.

7 Conclusion

We currently witness a growing interest in research on in-vehicle systems and their effects on drivers and driving performance. Inspired by previous research on in-vehicle interaction with touch-screen technologies, we compared different input and output techniques for in-vehicle systems to investigate and measure their effects on the driving activity and driver attention. As we stated in the introduction, driver attention is critical while drivers handle vehicles and interacts with emerging technologies in contemporary cars. We conducted an experiment with 32 subjects in four configurations to investigate the effects of input and output.

Our findings showed that when addressing in-vehicle systems design separating input and output made a difference. Using gesture input resulted in significantly fewer eye glances compared to touch input, but also inferior primary driving task performance and longer task completion times. Audio output caused the test subjects to make more longitudinal control errors compared to visual output, and had a significantly longer task completion time. Visual output, on the other hand, accounted for significantly more interaction errors and a drastically higher number of eye glances. Looking at the individual input/output configurations, our results showed that gesture/audio had the lowest number of eye glances, but also a longer task completion time and more longitudinal control errors than any of the other configurations.

Our results did not, on the other hand, indicate that fewer eye glances necessarily entails better primary driving task performance. On the contrary, audio output, which had fewest eye glances, seemed 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 load of the driver, distracting their cognitive attention from the primary task of driving the car. Further research might shed more light on this phenomenon.

Acknowledgements

We would like to thank all the participating test subjects in our experiment. Also, we want to thank several anonymous reviewers for comments on earlier drafts of this paper.

References

1. Alpern, M., & Minardo, K. (2003). Developing a Car Gesture Interface for Use as a Secondary Task. CHI 2003: New Horizons , pp. 932-933.
2. Bach, K. M., Jæger, M., Skov, M. B., & Thomassen, N. G. (2007). Interacting with In-Vehicle Information Systems: Understanding, Measuring, and Evaluating Attention. In Proceedings of the HCI 2009, ACM Press, pp. 453-462
3. Bach, K. M., Jæger, M., Skov, M. B., & Thomassen, N. G. (2008). You Can Touch, but You Can't Look: Interaction with In-vehicle Systems. Proceeding of CHI '08 , pp. 1139-1140.
4. Brooks, C., & Rakotonirainy, A. (2007) In-vehicle Technologies, Advanced Driver Assistance Systems and Driver Distraction: Research Challenges. In I. J. Faulks, M. Regan, M. Stevenson, J. Brown, A. Porter, & J. D. Irwin (Eds.), Distracted Driving (pp. 471-486). Sydney, NSW: Australasian College of Road Safety.
5. Brown, I. (1994). Driver fatigue in Human Factors, Vol. 36, Issue 2, 298-314. Human Factors and Ergonomics Society
6. Chewar, C. M., McCrickard, D. S., Ndiwalana, A., North, C., Pryor, J., & Tesselndorf, D. (2002). Secondary task display attributes: optimizing visualizations for cognitive task suitability and interference avoidance. Proceeding of the symposium on Data Visualisation (VisSum '02), Eurographics Association , pp. 165-171.
7. 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.
8. Eysenck, M. W. (2001). Principles of Cognitive Psychology (2nd ed.). Psychology Press
9. Geiger, M., Zobl, M., Bengler, K., & Lang, M. (2001). Intermodal differences in distraction effects while controlling automotive user interfaces. Proceedings Vol. 1: Usability Evaluation and Interface Design, HCI 2001 , pp. 236-267.
10. Gellatly, A. (1997) The Use of Speech Recognition Technology in Automotive Applications. Faculty of Virginia Polytechnic Institute and State.
11. Green, P. (1996). Customer Needs, New Technology, Human Factors, and Driver Science Research for Future Automobiles. Journal of the Society of Mechanical Engineers. University of Michigan Transportation Research Institute (UMTRI)
12. Green, P. (2001) Variations in Task Performance Between Younger and Older Drivers: UMTRI Research on Telematics in Association for the Advancement of Automotive Medicine Conference on Aging and Driving, Southfield, Michigan
13. Green, P. (2004) Driver Distraction, Telematics Design, and Workload Managers: Safety Issues and Solutions. University of Michigan Transportation Research Institute. SAE International.
14. Just, M. A., Keller, T. A., & Cynkar, J. (2008) A decrease in brain activation associated with driving when listening to someone speak. Carnegie Mellon University.
15. Kern, D., Schmidt, A., Arnsmann, J., Appelmann, T., Pararasasegaran, N., and Piepiera, B. (2009) Writing to your car: handwritten text input while driving. In Proceedings of Extended Abstracts on Human Factors in Computing Systems (CHI'09), ACM Press, pp. 4705-4710
16. Lansdown, T. C., Brooks-Carter, N., & Kersloot, T. (2004) Distraction from Multiple In-Vehicle Secondary Tasks: Vehicle Performance and Mental Workload Implications. Ergonomics , 47 (1), pp. 91-104.
17. Lee, J., Forlizzi, J., and Hudson, S. E. (2005) Studying the effectiveness of MOVE: a contextually optimized in-vehicle navigation system. Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI '05), Portland, Oregon, USA, ACM Press, pp. 571-580
18. MedicineNet. (2010). Retrieved from http://www.medicinenet.com/hemin_for_injection/glossary.htm
19. Nass, C., Jonsson, I., Harris, H., Reaves, B., Endo, J., Brave, S. and Takayama, L. (2005) Improving Automotive Safety by Pairing Driver Emotion and Car Voice Emotion. In

- Proceedings on Human Factors in Computing Systems (CHI'05), Portland, Oregon, USA. ACM Press, pp.
20. Pirhonen, A., Brewster, S., & Holguin, C. (2002). Gestural and Audio Metaphors as a Means of Control for Mobile Devices. *CHI Letters*, 4 (1), pp. 291-298.
 21. Rockwell, T. H. (1988) Spare visual capacity in driving - revisited: New empirical results for an old idea. In M. H. Freeman, P. Smith, A. G. Gale, S. P. Taylor, & C. M. Haslegrave (Eds.), *Vision in Vehicle II* (pp. 317-332). Elsevier Science.
 22. Sears, A. (1991) *Improving Touchscreen Keyboards: Design issues and comparison with other devices*. University of Maryland.
 23. Sears, A., Revis, D., Swatski, J., Crittenden, R., & Schneiderman, B. (1992) *Investigating Touchscreen Typing: The effect of keyboard size on typing speed*. University of Maryland.
 24. Strayer, D. L., & Drews, F. A. (2007) *Cell-Phone Induced Driver Distraction*. University of Utah.
 25. Tijerina, L. (2000). *Issues in the Evaluation of Driver Distraction Associated with In-Vehicle Information and Telecommunications Systems*. Transportation Research Center Inc
 26. Tsimhoni, O., Yoo, H. and Green, P. (1999) *Effects of Visual Demand and In-Vehicle Task Complexity on Driving and Task Performance as Assessed by Visual Occlusion*. University of Michigan Transportation Research Institute (UMTRI)