Abstract
To select the right modality for the interaction between drivers and the in-vehicle information system (IVIS) is crucial for safety reasons. This paper presents an experimental study to address this area. The study was carried out on a 160 degree car-driving simulation lab. There are 10 subjects participated in the experiment. We compared the subjects driving behavior on speech input/output only and speech input with speech+visual output interaction modalities with a simple IVIS. To judge the safety status of subjects’ driving performance, two independent variables which includes the average division of over speed and the average division of the car out of lane were measured as dangerous extent. Result indicates that it is not significant differences of driving performance by using synthetic speech to replace the visual display in the IVIS. It indicated that the visual presentation of a multi-modal IVIS can be acts as redundancy or complementary modality for auditory presentation, which will aids in relieving the resource demand.

1. Introduction
Today, a wide range of new in-vehicle technologies are being introduced into the car system. This includes Advanced Driver Assistance System (ADAS) and In-vehicle Information System (IVIS). Moreover, the in-vehicle use of portable computing devices (or nomad devices), such as mobile phone and GPS is increasing rapidly. Speech based multimodal interaction system for the IVIS is in the highest interests [1]. ADAS system is having the potential to enhance the safety on the road while the IVIS provide the services not directly relevant for the primary driving activities. As it requires the driver to have attention out of the road when performing with IVIS, it has the potential to increase the risk on the road. Many studies have pointed out that mobile phone use inside the car can cause high risk of traffic accident [2].

Safety issue is critical for implementing any system into the car. Beside all the technical questions, the problem regarding how to select the right modality for right information input and output is still remaining an issue. There is initial study done which compare different types of control for operating IVIS, incorporating radio, tape and CD functions with speech to traditional manual controls [3]. The driving performance was found to be significantly better while using voice controls than the manual controls. Subjectively, the voice controls with feedback were rated easier, most likeable, and most efficient to use while driving. However, some studies have shown that the interfaces using speech input-output in in-vehicle system may still present significant safety risks because of re-direction of attention if the concurrent tasks have a relatively high memory load or demands on the central executive function, in working memory [4, 5, 6].

Studies on investigating the distraction from the hands-free phone conversation showed that the hands-free was significantly more demanding than either the passenger conversation or in-vehicle tasks [7]. Significant level of visual distraction when the drivers proceeding through high volume intersections while performing demanding cognitive task by responding the difficult arithmetic questions through cell phone query. [8].

Thus the studies above revealed that the cognitive loading is a very important factor that determines the performance and safety of driving. To select the suitable modality for the interaction of the IVIS system is the critical issue for the success of the IVIS.

For drivers, to keep the eyes on the road is the essential for the safety on the road. It is normally prohibit to having any extra display inside the vehicle. But from the previously study on mobile phone using one can see that using speech as the only interaction modality can add heavy memory loading on the drivers and it also affect the driving behavior. In this paper, we will try to add visual display as the redundant information display recourse and study its effects to the driving performance.

2. Method
The experiment was carried out in a 160 degree driving simulator [10]. Each screen had a 1024*768 resolution. The graphics were rendered in real time in response to driver’s actions. No force feedback of tactile simulation was involved. The simulated car had an average performance and automatic gear shifting. The simulated environment consisted of a typical Swedish rural way with a speed limit of 90 km/h and standard road layout, delineation, signs and one meeting car in the opposing lane.

There are 10 subjects enrolled in the experiment. All of them are the students of Linköping University, with driving experience of 3.4 years. It is a within group experimental design.
2.1. Interface design and experimental process:

Visual presentation (Figure 1) was done on a partly covered CRT, which is located at the right hand side of the steering wheel (Figure 2).

![Figure 1: Screenshot of the in-vehicle information system](image1)

![Figure 2: The cockpit of driving simulator and the CRT](image2)

The primary task for the subjects to perform was driving a simulator car on a road with maximum speed of 90km/h. While driving, the subject was supposed to fulfill the task requirement by using the IVIS.

A simple protocol type of the IVIS with two functions, MP3 player and accommodation booking was implemented to conduct the experiment. Synthetic speech was generated from the Wizard-of-Oz approach whereby the experimenter manipulated the interface output respond based on the speech commands by driver. The synthetic voice was generated by Microsoft® Agent (it can be downloaded freely from [http://www.microsoft.com/products/msagent/default.asp](http://www.microsoft.com/products/msagent/default.asp)). The driving performance was recorded by the log file generated from the driving simulator and user interaction to the system was recorded by the log file generated from the IVIS. Time was used to match between the two log file. The vocabulary of the speech command is designed to be easy to learn, a dynamic synthetic speech based instruction of the interface is supplied to help the subject in the procedure. Training of driving the simulator car and using the IVIS was arranged before the test.

There were two test conditions in this experiment. One was asking the subjects to select songs for playing MP3. In the MP3 play task, subjects were asked to find three specific songs from the library of the IVIS, add them into the play list one by one and then reserve them together. The information of the price, address was presented by the synthetic speech, and the service supplied was presented by the visual based icon. To fulfill the task, the subjects were forced to use both spatial and audio attention to perceive enough information.

2.2. Independent variables:

Car speech and lateral position were measured as the driving behavior, and subjective workload was measured by using NASA-TLX questions. High dangerous extent indicates low safety status. Thus when the speed exceeds 95 km per hour, it is considered having high dangerous extent. Similar for lane position, when the subjects drove beyond either the centre line or the kerbed, it is considered having dangerous extent. For this experiment, the simulator road is based on Swedish standard road of 3.3 meters wide and the simulated car used is Volkswagen Golf. NASA TLX was therefore introduced to measure it in this experiment as the third independent variable. Diagram shown in Figure 3 illustrate the safety margin defined for this experiment.

![Figure 3: Safety margin defined for lateral position and speed](image3)

![Figure 4: Integration of over speed alone the over speed time](image4)
The average division of over speed measured in this experiment is the integration of the over speed \( \left( v_o \right) \) along the over speed time \( \left( t_o \right) \) divides the total task performance time. (See Figure 4 and equation (1))

\[
R_{\text{over-speed}} = \frac{\int_0^T v_o \, dt_o + \int_0^T v_o \, dt_o + \ldots + \int_0^T v_o \, dt_o}{T_{\text{total}}} \tag{1}
\]

Another variable, the average division for the car out of lane was defined as the integration of the distance for the left ties pass through the middle line \( (d_l) \) along the passing time and the integration of the distance for the right ties pass through the right edge \( (d_r) \) of the road along the passing time. (See the Figure 3, Figure 5 and equation (2)).

\[
R_{\text{dangerous-lateral-position}} = \frac{\left( \int_0^T d_l \, dt_l + \int_0^T d_l \, dt_l + \ldots + \int_0^T d_l \, dt_l \right)}{T_{\text{total}}} \tag{2}
\]

\[
+ \left( \int_0^T d_r \, dt_r + \int_0^T d_r \, dt_r + \ldots + \int_0^T d_r \, dt_r \right)
\]

3. Result:

3.1. Measurement of speed

Paired t-test was used for the analysis. The average division of over speed, is 0.78 \((\pm 0.37)\) Km/hr of MP3 and 1.26 \((\pm 0.22)\) m of accommodation booking. In addition, \( t\text{-Stat}= -1.1 \) (DF=9, \( \alpha = 0.05 \)), which is less than the \( t\text{-Crit}= 2.26 \), thus there is no significant difference of the two treatments (Figure 6).

\[
R_{\text{dangerous-lateral-position}} = \frac{\left( \int_0^T d_l \, dt_l + \int_0^T d_l \, dt_l + \ldots + \int_0^T d_l \, dt_l \right)}{T_{\text{total}}} \tag{2}
\]

3.2. Latency measurement

The average division of the car out of the lane is 0.78 \((\pm 0.37)\) m of MP3 and 1.26 \((\pm 0.22)\) m of accommodation booking (Figure 3). In addition, \( t\text{-Stat}= 0.79 \) (DF=9, \( \alpha = 0.05 \)), which is less than the \( t\text{-Crit}= 2.26 \), thus there is no significant difference of the two treatments (Figure 7).

3.3. Mental Workload

For the total scores of NASA TLX, \( t\text{-Stat}= -2.8 \) (DF=9, \( \alpha = 0.05 \)), which is less than the \( t\text{-Crit}= -2.26 \), thus there is significant difference between the two mean of the two treatments.

4. Discussion and Conclusion

It is crucial factor to keep the driver safe and other road users safe when any IVIS is introduced into the car cockpit. Many researches regarding IVIS are focusing on the technical itself and how to make the system working. One of the most difficult question for a multimodal interactive system design is how to choice the right modality for the right information, at right timing. When regarding the interactive system design, interaction itself with the system is only part of the consideration. The total interactions, happened between the drivers to the IVIS and to the road shall always consider together. The result of this study is mean to have certain guidance for speech based multimodal interaction for IVIS.

Answering hand held mobile phone is prohibited by law in a lot of country because it is supposed to increase the traffic...
accident rate. However, the truth is if we can’t totally prohibit driver using mobile phone when they are driving, the problem is still there because by using hand free mobile phone in the car cockpit in stead may also affect the driving performance. That is also the reason of multimodal IVIS is selected to be studied in this paper. The interaction with the in-car computer may affect the driving performance of the driver, but if we can’t stop driver of using the IVIS, then the most important thing is to judge whether the driving behavior is still in safety status when the driver is interact with the IVIS and which interaction modal is the best solution for driver to keep their driving performance in the safety range. Thus, two independent variables which include the average division of over speed and the average division of the car out of lane were measured as dangerous extent to judge the safety status of subjects’ driving performance in this experiment.

In addition, when the traditional visual display was replaced by the synthetic speech display in the IVIS, the spatial attention of driver released successfully from the screen so that it can concentrate on the road. It seems to be helpful to improve the safety of driving behavior based on the information processing modal. However, from the statistic analysis of the data of the average division of over speed, there is no significant effect between the two treatments; the data of the average division of the car out of lane also indicates no significant difference. There are many deduction which may contribute to this none significant result. The speech only interface of the IVISs may impose substantial cognitive load because of its potential problems, such as the additional cognitive processing cost with the comprehension of synthetic speech [3] and dialog grammar, understanding the menu structure and recalling the speech command. For the car driving task, the noise in the cockpit which may come from the car and the environment is also a possible cause. All these disadvantages of the speech based interaction may affect the driving performance. On the other hand, human beings may adjust their own behavior to adapt to the different working environment. For example, when they are performing the accommodation book task which requires them to check the service icon displayed on the screen while driving, they can try to avoid watching the monitor when the car in the curvature. This kind of adaptation may also contribute to the none significant result. Thus, synthetic speech based interaction of in-car computer is not as helpful as it supposed to be in this experiment.

The performance results in this paper did not show a significant effect in terms of deviation of speed and lane position between tasks MP3 and Accommodation Booking. By comparing the two tasks, accommodation book is much more complicated and required higher mental demand than select songs on MP3. It is not surprise that a significant amount of mental workload especially from the mental and physical demand was observed from task Accommodation Booking using the NASA TLX. When the workload increases from task MP3 to Accommodation Booking, there is no significant effect of driving between the two tasks. The data from the deviation of lane position indicates no significant difference from the statistic analysis of the data of the deviation of speed. One possibility might be that the subjects deal with such higher task demands by investing more effort, performing behavioral adaptation, changing working strategy or neglecting the subsidiary information to adapt to the difficult working environment. It indicated that the visual presentation of a multi-modal IVIS can be acts as redundancy or complementary modality for auditory presentation, which will aids in relieving the resource demand.

5. Reference:


