FFI

SIMSI - Safe In-Vehicle Multimodal Speech Interfaces

Project within Vehicle and Traffic Safety

Authors: Per Karlsson, Mecel AB; Staffan Larsson, University of Gothenburg ; Fredrik Kronlid, Talkamatic AB

Date: 2014-03-15

FFI

Content

FFI in short

FFI is a partnership between the Swedish government and automotive industry for joint funding of research, innovation and development concentrating on Climate & Environment and Safety. FFI has R&D activities worth approx. €100 million per year, of which half is governmental funding. The background to the investment is that development within road transportation and Swedish automotive industry has big impact for growth. FFI will contribute to the following main goals: Reducing the environmental impact of transport, reducing the number killed and injured in traffic and Strengthening international competitiveness. Currently there are five collaboration programs: **Vehicle Development, Transport Efficiency, Vehicle and Traffic Safety, Energy & Environment and Sustainable Production Technology.**

For more information: www.vinnova.se/ffi

1. Executive summary

Driver distraction is one common cause of accidents, and is often caused by the driver interacting with technologies such as mobile phones, media players or navigation systems. In SIMSI, we have taken major steps towards developing a system which enables safe interaction with technologies in vehicles, by reducing the cognitive load imposed by the interaction and minimizing head-down time.

The primary goal of the project has been to carry out research focusing on reducing driver distraction using integrated multimodality and dialogue strategies for cognitive load management. Based on the research, the project has developed interaction strategies for minimizing distraction, and empirically investigated different interaction strategies from a safety perspective. To reach this goal, a number of activities have been carried out:

- First, a technical setup for testing and demonstrating the system was constructed to enable simulator tests and demonstrations of the system.
- Second, a couple of applications to interact with were designed and implemented, to enable testing of the generic interaction strategies developed in the project.
- Interaction strategies for reducing visual and cognitive distraction were designed and implemented. The interaction strategies were divided into (1) multimodal solutions to reduce head-down time, and (2) solutions for reducing the cognitive load of the driver.
- Finally, tests of the implemented applications and strategies were carried out that potentially allows results to feed back into the development cycle.

SIMSI has developed existing technology combining integrated multimodality designed to minimize visual distraction with dialogue strategies for cognitive load management. SIMSI has brought this technology closer to market by making clearer how the Talkamatic Dialogue Manager (TDM) integrates with existing commercial technologies. We have created an integration guide for TDM, and accumulated experience in integration with commercial-strength HMI. We have also developed a testing platform for distraction in in-vehicle interaction which will feed future research and development. The University of Gothenburg aim to re-use the platform in future projects, including EU projects.

Although test results are still being analyzed, we believe that the SIMSI system will be shown to reduce distraction, cognitive load and head-down time considerably when compared to other state-of-the-art in-vehicle interaction models.

2. Background

Driver distraction and safety

Driver distraction is one common cause of accidents, and is often caused by the driver interacting with technologies such as mobile phones, media players or navigation systems. The so-called 100-car study (Neale et al., 2005) revealed that secondary task distraction is the largest cause of driver inattention, and that the handling of wireless devices is the most common secondary task. The goal of the research proposed here is to design systems which enable safe interaction with technologies in vehicles, by reducing the cognitive load imposed by the interaction and minimizing head-down time. Such systems will contain both active components, which track and adapt to the driver's cognitive load, and passive systems, which by their design enable interactions imposing minimal cognitive load on the driver.

Multimodal Dialogue Systems

In the vehicle industry, we talk about Human Machine Interfaces (HMI's) or Driver-Vehicle interfaces, some of which include speech. In academia, many researchers instead talk about *Multimodal Dialogue Systems* (or MDS's for short). A multimodal dialogue system enable spoken communication between humans and machines, but complements the spoken modality with traditional human-machine interaction modalities such as visual output (screen, head-up display) and haptic input (scroll wheels, buttons, etc.). The research proposed here aims to further develop an existing MDS by combining a particular kind of multimodality (*integrated* multimodality) with dialogue strategies for cognitive load management.

The Talkamatic Dialogue Manager

A dialogue manager is the central component of a dialogue system. It keeps track of the dialogue context, including dialogue history and dialogue plans, and uses this information to understand user utterances and select own utterances. Based on academic research on dialogue systems (Larsson 2002 and later work), Talkamatic AB has developed the Talkamatic Dialogue Manager (TDM) with the goal of being the most competent and usable dialogue manager on the market. TDM provides a general interaction model founded in human interaction patterns, resulting in a high degree of naturalness and flexibility which increases usability.

TDM deals with several interaction patterns which are basic to human-human linguistic interaction, and offers truly integrated multimodality which allows user to freely switch between modalities. All these solutions are domain-independent which means that they need not be implemented in each application. Using Talkamatic technology, dialogue behaviour can be altered without touching application properties, and can be updated without touching the dialogue logic. This makes testing of different dialogue strategies, prompts etc. considerably quicker and easier than when using regular state-machine-based dialogue systems.

In addition, as the dialogue strategy is separated from the application logic, development time for new dialogue applications can be significantly reduced. Furthermore, the developer designing the application does not need to be a dialogue expert as the dialogue design is built into the dialogue manager. The Talkamatic dialogue manager architecture is also well-suited for an Appstore vehicle scenario, where the user can download extra functionality for the car's infotainment system.

Integrated multimodality in TDM

There are reasons to believe that multi-modal interaction is more efficient and less distracting than uni-modal interaction (Oviatt et. al. 2004). TDM supports multi-modal interaction where voice output and input (VUI) is combined with a traditional menu-based GUI with graphical output and haptic input. In cases where a GUI already exists, TDM can replace the GUI-internal interaction engine, thus adding speech while keeping the original GUI design. All system output is realized both verbally and graphically, and the user can switch freely between uni-modal (voice or screen/keys) and multi-modal interaction.

To facilitate the browsing of lists (a well known interaction problem for dialogue systems), Talkamatic has developed its Voice-Cursor technology (Patent Pending). It allows a user to browse a list in a multi-modal dialogue system without looking at a screen and without being exposed to large chunks of readout information.

A crucial property of TDM's integrated multimodality is the fact that it enables the driver of a vehicle to carry out all interactions without ever looking at the screen, either by speaking to the system, by providing haptic input, or by combining the two. We are not aware of any current multimodal in-vehicle dialogue system offering this capability.

Additional information is available at www.talkamatic.se

Mecel Populus

The Mecel Populus suite is a complete tool chain for designing, developing and deploying user interfaces for distributed embedded systems. It minimizes the time and cost of producing eye-catching, full-featured HMIs (Human Machine Interfaces).

The Mecel Populus concept has several unique features compared to traditional HMI development. These features, when combined, remove the barriers that traditionally exist between the people working with requirements, system engineering, HMI design and implementation. Mecel Populus has been designed for the automotive industry to deliver high performance user interfaces with short time-to-market and to enable efficient software life cycle management. Additional information is available at www.mecel.se/products

3. Objective

The primary goal of the project has been to carry out research focusing on reducing driver distraction using integrated multimodality and dialogue strategies for cognitive load management. Based on the research, the project has developed interaction strategies for minimizing distraction, and empirically investigated different interaction strategies from a safety perspective.

To reach this goal, a number of activities have been carried out:

- First, a technical setup for testing and demonstrating the system was constructed to enable simulator tests and demonstrations of the system.
- Second, a couple of applications to interact with were designed and implemented, to enable testing of the generic interaction strategies developed in the project.
- Interaction strategies for reducing visual and cognitive distraction were designed and implemented. The interaction strategies were divided into (1) multimodal solutions to reduce head-down time, and (2) solutions for reducing the cognitive load of the driver.
- Finally, tests of the implemented applications and strategies were carried out that potentially allows results to feed back into the development cycle.

4. Project realization

4.1 WP1: Integration

WP1 has been by far the most challenging part of the project, and has taken much more resources than expected. The "peerness" of the two engines motivated a symmetric relation between the two engines, rather than a strict master-slave relation. Both engines keep track of the current state of the interaction and manage transitions between states resulting from user or system actions. To this end, a lock mechanism was introduced, where one of the engines was granted temporary control. While owning the lock, the engine performs changes to its own state and informs its peer engine about every such change. This way, each engine is allowed to have its own model of the interaction state, just as when running without its "peer", while the integration protocol allows these states to stay in sync with each other.

Throughout the implementation work in WP1, this general and simple solution has been gradually extended to cover more advanced scenarios. For instance, TDM deals with uncertainty related to weak speech recognitions, with tentative assumptions that can be withdrawn or kept depending on subsequent user utterances. In contrast, Populus has no need for this functionality.

All of the integration issues were eventually dealt with successfully. For Talkamatic this experience is invaluable, as it demonstrated the differences between TDM and other interaction engines, and how such differences can be bridged.

4.2 WP2: Application development

The application suite was quite limited. It consisted of a voice-enabled standard telephone application, accessible from a main menu. The main menu also contained icons for music and messaging applications, although without content. There was also support in the platform for a mocked navigation application which simulated turn-by-turn navigation instructions.

The phone application allowed the user to browse a contact list where some entries have a single number while others have multiple numbers. The user can use voice, touchscreen or buttons on the steering wheel to select a number. When the number is called, the user is presented with a screen indicating that a call is currently active.

Everything that can be done using the haptic UI can be done using voice (calling, selecting, hanging up, accepting an incoming call etc.). The user can also use one shot utterances, e.g. “Call Alexander’s mobile number”. The voice browsing functionality means that an item which is currently in focus in the GUI is read out to the user.

4.3 WP3: Integrated multimodality for reduced head-down time

This WP focused on evaluation of the voice cursor concept. We here describe the tests carried out in WP3. Data from the tests are still being processed and the final results will be described in the After-report.
The main point of the visual distraction tests is to investigate how the “eyes-on-road” time during interaction varies between different modality conditions. Our main interest was on the contribution of the voice cursor technology.

The following four conditions were tested:

- GUI only (haptic only in, graphics only out)
- GUI with voice cursor (haptic only in, graphics and speech out)
- Multimodal with voice cursor (haptic and speech in, graphics and speech out)
- Speech-only with voice cursor (haptic and speech in, speech only out)

For each condition, we used two difficulty levels: easy and difficult. For both levels, the task is to drive along a softly curving road while keeping distance to one car in front of you and one car behind you. In the easy condition, the other cars have a constant speed. In the difficult condition, the other cars are speeding up and braking erratically, and the car behind you may indicate (by honking its horn) that you're going too slowly.

This way of testing, which we informally refer to as the “annoying cars” setup, differs from existing experimental setups such as the ConTRe task (Engonopoulos et al, 2008). In the latter, the driver tries to match two vertical lines representing the vehicle's position and the target (reference) position. Our setup has the advantage of being more realistic, although we acknowledge that it is still far from driving in real traffic. Initial tests will be carried out to verify the adequacy if the ``annoying cars'' setup for our purposes.

The application used in the tests has basic phone functionality: browsing a list of contacts, and calling people up. At regular intervals, the driver receives a spoken instruction (with a voice different from the dialogue system), e.g. “You just remembered you need to call up Ashley on her mobile number.” The user should then carry out this instruction as efficiently and completely as possible.

Questionnaires were used to collect data (1) before the tests, (2) after each modality condition, and (3) after the whole test session. The pre-test questionnaire collects background data about the test subject, including driving experience, familiarity with technologies such as voice control and GPS systems. After each modality condition, the test subject is asked to rate (on a scale from 1 to 5 where 1 represents full disagreement and 5 full agreement) their agreement with 12 statements concerning their experience with the interaction.

4.4 WP4: Dialogue strategies for reduced cognitive load

The objective of this work package was to find out what dialogue behaviours best serve to minimize the driver's cognitive load from three different aspects:

- Interruption and resumption strategies: When and how should dialogue be interrupted and resumed to minimize distraction? (WP4.1)
- Rhetorical strategies: Can rhetorical strategies designed to make it easy to understand and accept information also be used to decrease distraction? (WP4.2)
- Pausing and turn-taking strategies: Can within-turn pause behaviour be manipulated to minimize distraction? (WP4.3)

Earlier research indicates that spoken communication may affect driver behaviour negatively. However, these investigations are typically based on very artificial types of communication, where the driver's dialogue partner pays no attention to the driver's distraction level or the traffic situation. By contrast, other studies have shown that when the driver talks to a passenger in the car or uses a mobile phone for normal everyday conversation, this has virtually no adverse effects on driving behaviour (Esbjörnsson et al 2006). The reason is that the passenger continually adapts his communication to minimize the cognitive load of the driver, for example by keeping quiet when the driver is focused on the driving task.

Work in this WP has concentrated on specifying the strategies to be tested, implementing them on the SIMSI test platform, and on design of the tests. We will here describe the planned tests for WP4.

WP4.1: Interruption and resumption strategies

The effects of interruption and resumption strategies on cognitive distraction will be tested using interactions with the SIMSI system. Cognitive load is continuously measured, but does not itself affect the behaviour of the system. Instead the "annoying cars" concept described in Section 2.3 is use to induce a stressful driving situation, and instructions displaying different pausing behaviours are played during the stressful periods.

The two behaviours which will be compared are the following:

1. Continue talking under difficult condition
2. Stop talking during difficult condition; resume interaction when difficult condition ends

To ensure that there is an ongoing interaction and that it is paused under the difficult condition, the switch in condition is actually controlled from the dialogue system so that the difficult condition will be triggered at a certain point in the interaction. This means that no actual cognitive load detector is needed in the experiments. Instead, the idea is to use data collected during these experiments as the starting point for building and automatically training a cognitive load detector module.

WP4.2: Rhetorical strategies

The effects of rhetorical strategies on cognitive distraction will be tested using pre-recorded prompts which provide navigation instructions while driving. The user does not have to respond to these commands. Cognitive load is continuously measured, but does not itself affect the generation of pauses.

WP4.3: Pausing and turn-taking strategies

As in WP4.2, the effects of pausing on cognitive distraction will be tested using pre-recorded prompts which provide navigation instructions while driving. The user does not have to respond to these commands. Cognitive load is continuously measured, but does not itself affect the generation of pauses. Instead, and as in WP4.1, the "annoying cars" concept described in Section 2.3 is use to induce a stressful driving situation, and instructions displaying different pausing behaviours are played during the stressful periods.

FFI

WP5: User testing and evaluation

One goal of SIMSI is to conduct ecologically valid test of the applications, and to let the results of these tests feed back into the development of the system. Basically, we want to find the best interaction solutions and to verify these experimentally, especially in cases where it is not intuitively clear what is best. This involves implementing variants of a behavior, testing them on naive users, collecting data from these interactions, and establishing statistically significant results based on the collected data.

As part of the preparations for the experiments in WP4, GU spent a considerable amount of effort on setting up a complete testing environment based around a driving simulator and various kinds of biophysical measurement devices.

Using internal funding, GU acquired the SCANER driving simulator software and various pieces of hardware (steering wheel, driver’s chair with pedals and gear stick). We decided to go for a more realistic simulator than used in academic research in in-vehicle interaction, for reasons of ecological validity.

To measure where the driver is looking, we acquired (also using internal funding) the SmartEye Pro eye tracker. One important advantage of this eyetracker is that it allows head movements without data loss.

Finally, we added the CStress stress measurement device, which records heart rate, skin conductance and optionally also breathing. It proved quite difficult to find the appropriate kind of equipment, but it appears that several similar products aimed for a more general public are close to commercial release.

In addition to the above-mentioned equipment, the SIMSI dialogue system was connected to allow interaction with it using a microphone as well as buttons on the steering wheel for input and speakers and a display (an Android tablet) for output.

A video camera placed behind and to the right of the driver recorded all interactions (sound and video). Furthermore, all user utterances were recorded using the microphone which was used for interaction.

Scripts were written to collect and synchronize data from the driving simulator, the eyetracker, the dialogue system and the other parts of the system. Another script was written to supervise the test sessions.

5. Results and deliverables

5.1 Delivery to FFI-goals

- A demonstrator for multimodal HMI which is arguably superior to any in-vehicle interaction design on the market with respect to safety and usability when operating a vehicle.
- Increased knowledge about multimodal dialogue strategies for minimizing cognitive load and head-down time in in-vehicle interaction. Although the scientific results are not yet complete (they will be included in the follow-up report), we believe that we will also be able to provide significant test results demonstrating that the dialogue strategies in SIMSI decrease distraction and increase safety.
- *Support environments for innovation and collaboration,* by supporting the Centre for Language Technology at Göteborg University in its ambitions to become a hub for research on in-vehicle dialogue, and in general by strengthening the relations between academia and industry in the area of safe interaction in vehicles in the VG region.
- *Strive to ensure that new knowledge is developed and implemented, and that existing knowledge is implemented in industrial applications*, by further developing the TDM system and implementing it in demonstration and testing environments, thereby working towards ensuring its implementation in industrial in-vehicle applications

5.2 An integrated demonstration and testing platform

The System Integration (WP1) work package is the fundamental work package in the project, aiming to deliver the research system. The research system (the SIMSI system) consists of integration between Talkamatic's TDM, a dialogue manager for spoken dialogue, and Mecel Populus, software for specifying, verifying and deploying car HMIs. (This section explains the SIMSI research system. The research and testing platform used in SIMSI is described in below in Section 5.5.)

The major challenge in integrating Populus and TDM is that both systems keep track of the current state of the interaction and manage transitions between states resulting from user or system actions. Hence, there is a need to keep the systems in sync at all times. This is managed by a Transition Queue (TQ) module which keeps a lock which can be grabbed by either system at any time, unless it has already been grabbed by the other system. The systems then enter into a master-slave relation where the master is the system which owns the lock. The master tells the slave how the interaction state is to be updated, and the slave only waits for messages from the master until the lock has been returned to the TQ.

The synchronization is also complicated by the inherently different interaction for a user between a voice driven and a graphic driven environment. In the voice environment questions are asked and answered. Each interaction has a timespan. For example showing the phone numbers of a contact requires the engine to read them out, while the display can show all numbers at the same time. These timing and interactions have to be addressed when building both the graphical and voice interface.

5.4 Integrated multimodality for reduced head-down time

We have collected data from 20 test subjects interacting with the dialogue system. The eyetracker equipment was used for capturing where the driver is looking. Apart from collecting data about whether the driver is focusing on the driving simulator screens (eyes-on-road) or on the GUI, we also recorded fine-grained eye tracker data showing where the driver was looking at any given moment.

We hypothesise that in the GUI-only condition, there will be less eyes-on-road time than in the other three conditions, since the driver does not have to look at the screen in order to complete the task. Apart from testing this hypothesis, we are generally interested in which condition(s) gives the best results with respect to eyes-on-road time, task success, task completion time and usability (rated subjectively using a questionnaire. For example:

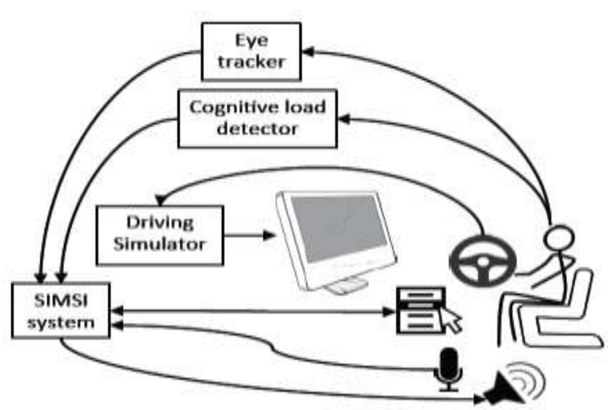
- How much does the Voice Cursor help minimize distraction on its own, without the speech input modality? Comparing this condition to the GUI-only condition will help us understand and quantify the contribution of the voice cursor to minimizing visual distraction, e.g. when browsing lists.
- Which modalities are preferred by users when all modalities are available all the time? How much (if at all) is visual distraction, driving quality and usability improved in this condition compared to (1) GUI only and (2) GUI with Voice Cursor?
- How will removing the visual output modality affect visual distraction, driving quality and subjective usability? Is it a good idea to turn off the screen in some circumstances, e.g. under difficult driving conditions, to eliminate visual distraction, or will usability suffer?

5.5 Dialogue strategies for reduced cognitive load

The test will be carried out in April 2014 and will be described in the after report.

5.6 Methodology for user testing and evaluation

The test environment consists of two parts, apart from the dialogue system: a driving simulator (SCANeR from Oktal) and an eye tracker (Smart Eye Pro from Smarteye), and instruments for measuring cognitive load (CStress).



In our setup we have three monitors, giving the user a wide field of view. We also have a gaming steering wheel, including pedals, gear lever and a driver's seat. These are used mainly to control the driving simulator, but there are also a number of buttons on the steering wheel which are used to browse the menus in the HMI and as Push-to-talk (PTT). An Android tablet showing the HMI GUI is placed in front of the user, trying to match the position of a display in a car. Both TDM and Populus run on the same desktop computer as the driving simulator, and a Populus Android app runs on the tablet. The app allows the user to select items by tapping them, as well as scrolling in lists in normal smart phone fashion. The eye tracker runs on a separate desktop computer, as it requires a substantial amount of processing power.

Figure 1: SIMSI test environment in action

Studio software that comes with the driving simulator is used to design and run scenarios. The scenarios govern how autonomous traffic should behave and events, such as weather change and the state of traffic signals. The simulator logs data for the environment and each vehicle. Data like lane deviation and how the user handles instruments, e.g. steering wheel and pedals can be used to measure cognitive load. At a later stage this kind of data can also be used to trigger behaviour in the dialogue system.

We believe that the SIMSI test setup is quite unique in a university setting, and it will be an invaluable resource in future works on in-vehicle dialogue systems.

6. Dissemination and publications

6.1 Knowledge and results dissemination

A recent development relevant to SIMSI is the adaptation of mobile devices to the in-vehicle environment, exemplified by Apple's CarPlay (https://www.apple.com/ios/carplay/) as well as the Open Automotive Alliance (http://www.openautoalliance.net) which aims to bring the Android OS to cars. There are concerns in the industry about driver distraction and safety risks associated with bringing apps designed for mobile devices into the vehicle environment, and recognition of the need to create safe voice-based user interfaces.

Talkamatic are currently taking part in three EU-funded projects where TDM is used and where SIMSI results feeds into research and development. Of particular relevance is the project **"SIMPLI-CITY – The Road User Information System of the Future"** (http://simpli-city.eu/) which aims to foster the usage of full-fledged road user information systems.

Results of the SIMSI project, both research results and the improvements and further development of the SIMSI dialogue system and the Talkamatic dialogue manager, has a clear potential to fill an important role in future development of safe in-vehicle dialogue systems.

6.2 Publications

Larsson, S. ; Berlin, S. ; Eliasson, A. et al. (2013). Integration and test environment for an in-vehicle dialogue system in the SIMSI project, *Proceedings of the SIGDIAL 2013 Conference*. s. 354-356. ISBN/ISSN: 978-1-937284-95-4

Larsson, S. ; Berlin, S. ; Eliasson, A. et al. (2013). Visual distraction test setup for an multimodal in-vehicle dialogue system. *Proceedings of the Workshop on the Semantics and Pragmatics of Dialogue*. 17 s. 215-217.

We expect several more publications based on the results of the WP3 and WP4 user tests.

7. Conclusions and future research

The main conclusions of SIMSI are the following:

- We have brought TDM closer to market by making clearer how TDM integrates with existing commercial technologies. We have created an integration guide for TDM, and accumulated experience in integration with commercial-strength HMI.
- We have developed a testing platform for distraction in in-vehicle interaction which will feed future research and development. GU aim to re-use the platform in future projects, including EU projects.
- Preliminary test results indicate that multimodality decreases visual distraction compared to traditional HMI, and more specifically that TDM's Voice Cursor technology further reduces distraction and increases user satisfaction.

Future research includes

- Developing the VoiceCursor concept further to handle fast scrolling and other kinds of haptic input such as gestures
- Improving the test platform further to simplify the test procedures and the post-processing and analysis of collected data
- Carrying out more tests, if possible including tests in real traffic, to further investigate the contribution to decreased distraction from variants of the dialogue strategies tested in SIMSI. We also want to carry out longitudinal tests where users have access to the system over a longer period of time

8. Participating parties and contact person

Staffan Larsson
Professor of Computational Linguistics
Department of Philosophy, Linguistics and Theory of Science
University of Gothenburg
Email: sl@ling.gu.se
Telephone: +46(0)31 7864378

Per Karlsson
Project Manager
Mecel AB
Email: per.karlsson@mecel.se
Telephone: +46(0)31 720 4526

Fredrik Kronlid
Managing Director
Talkamatic AB
Email: fredrik@talkamatic.se
Telephone: +46(0)703 60 21 90