

# RIVER

## Reduced Distraction via Voice Interaction



### FFI - Fordonsstrategisk Forskning och Innovation

Filip Frumerie, Volvo GTT

Ida Esberg, Volvo GTT

Christian Jernberg, Volvo GTT

Annika Silvervarg, LiU

Arne Jönsson, LiU

Sofia Lindvall, LiU

Jonatan Andersson, LiU

Göteborg, 2016-09-20





## Content

<b>1. Executive summary</b>	<b>3</b>
<b>2. Background</b>	<b>4</b>
<b>3. Objective</b>	<b>6</b>
<b>4. Project realization</b>	<b>7</b>
4.1 State of the Art	7
4.2 User needs	8
4.3 How should speech interfaces be designed and implemented?	8
4.4 Voice interaction demonstrator	9
4.3 Methodology and evaluation	9
<b>5. Results and deliverables</b>	<b>11</b>
5.1 Delivery to FFI-goals	11
<b>6. Dissemination and publications</b>	<b>12</b>
6.1 Knowledge and results dissemination	12
6.2 Publications	12
<b>7. Conclusions and future research</b>	<b>13</b>
<b>8. Participating parties and contact person</b>	<b>14</b>
<b>9. References</b>	<b>15</b>

### 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](http://www.vinnova.se/ffi)

## 1. Executive summary

To allow drivers to interact with voice is seen as an important countermeasure to unsafe driving since voice technology aims to reduce the visual-manual load, allowing the driver to keep his/her eyes on the forward road scene. Studies have shown an increased risk when using visual-manual interaction compared to using voice control while interacting with nomadic devices. This would support the idea to increase the usage and integration of voice interaction in vehicles. However, at the same time voice interaction may still impose a high level of cognitive distraction. RIVER aims to further explore the relationship between visual and cognitive load and voice interaction. New and improved means for the driver to interact safely with the vehicle via voice will be developed and demonstrated in a demonstrator that will be evaluated in a real field setting with professional drivers. In addition to safety, the evaluation will also include acceptance and efficiency of the interaction.

## 2. Background

New and improved means for the driver to interact safely with the vehicle is one of the important pre-requisites to reach the high safety goals and visions for the Volvo Group. To allow the driver to interact with voice has been seen as one important countermeasure to unsafe driving but it is clear that a large research gap still exist both when it comes to the technical maturity of speech technology, fast changing user needs, evaluation methodology targeting speech interfaces as well as how the actual interaction is designed.

### **Nomadic devices, ITS applications and connected vehicles**

We see a global increased use of mobile phones and other nomadic devices as well as more connected vehicle applications such as music/audio playing, smartphone apps, navigation, roadside assistance, contextual help/offers, parking apps etc.

Connected navigation is a key trend and there are several examples like Nokia HERE Auto that offers cloud connected navigation service and also Apple with their CarPlay where you can connect your iPhone to the cars infotainment system.

In general brought-in smartphone integrated solutions are reaching high levels of availability in the European market where over 60% of the OEM offer such solutions (Herrera, 2014). In addition to the market pull from consumers bringing in more and more advance nomadic devices additional C-ITS functions are being offered, e.g. improved traffic jam information and Road Work Warnings (MoU Car 2 Car, 2011).

### **Voice control and speech interfaces**

Through voice interaction, drivers are given a chance to spend more time looking on the road and less time engaging in visual-manually interaction (e.g. trying to localize the correct buttons/stalks and navigate correctly through menus). Voice interfaces can also allow for short-cuts when navigating through menus. Instead of step-by-step navigating to the wanted item the driver would not have to look away from the road at a visually displayed menu when using voice interaction. However, voice interaction may still impose a relatively high level of cognitive distraction.

While cognitively distracting tasks have been shown to affect different types of driver behavior in different ways, in particular responses to brake light onsets and artificial stimuli such as the Detection Response Task (see Engström, 2011, for a review), it is unclear to what extent cognitive distraction actually leads to an increased crash risk. Existing studies of crashes and near crashes have consistently demonstrated negative effects of visual distraction such as looking away from the forward road scene for example when dialing or texting on a mobile phone (Klauer et al., 2006; Olson et al., 2009; Hickman et al., 2010; Victor et al., 2014). In all these studies, visual-manual tasks were associated with the highest risk while non-visual but cognitively distracting tasks such as talking/listening on a mobile phone or a CB radio did not significantly increase risk. On the contrary, in several of these studies (Olson et al., 2006; Hickman et al., 2010; Victor et al., 2014), crash/near crash risk was significantly reduced during cognitively distracting tasks. In Victor et al. (2014), focusing specifically on rear-end crashes/near crashes, the risk while engaged in talking/listening with a cell phone was reduced by 10



times relative to baseline. These data suggests that voice controlled interfaces could offer a safe mode of interaction with in-vehicle devices. However, it should be pointed out that voice technologies have not yet been evaluated in naturalistic data (due to their limited deployment).

Based on this, it is important that in the current project RIVER further investigate the relationship between voice based interfaces and visual *and* cognitive distraction.

Key starting points will be existing or draft ISO standards for measuring of visual and cognitive demand, such as the measurement of driver visual behavior based on eye movements (ISO 15007), visual occlusion (ISO 16673) and the Detection Response Task (DRT; ISO CD 17488). The on-going efforts by NHTSA to develop specific HMI guidelines for auditory-vocal interfaces (Phase 3 NHTSA guidelines, see e.g. US DOT, 2014), focusing on the DRT methodology, will also be closely followed.

### 3. Objective

The general objective of this project is to *assess, develop and demonstrate new improved voice functionality* with the goal to *improve safety and efficiency* and at the same time create a *high user-acceptance*. The project is directly related to the following target specified within the FFI program:

**A: Vehicle and Safety analysis** where the focus is to evaluate safe driver-vehicle interaction in real field setting with professional drivers focusing on distraction, acceptance and efficiency of interaction.

- How do truck drivers communicate today (e.g. what functions and with what tools – brought-in, build-in) and what speech interaction needs could they possibly foresee?
- What features in trucks would drivers prefer to control by voice, additional to the ones already controllable by voice?
- Is interaction via voice indeed a safer and more efficient way to communicate compared to traditional interfaces?
- How can current evaluation methods and tools for voice evaluation be improved?

**D: Driver support and related interface between driver and vehicle** where the focus is on new and improved means for the driver to interact safely with the vehicle via voice.

- How can Volvo's voice control interaction be made more efficient? E.g. non-menu based system, more natural speech.
- How can apps and other nomadic device related features be safely integrated via voice control system run in the vehicle?
- Can tell tale warnings be better presented to the driver by using voice control? If yes, how should the interaction be designed?

## 4. Project realization

The project was composed of four work packages (WP). WP1 was dedicated to planning, division of responsibilities and management of the other work packages. WP2 and WP4 provided the user needs, the methodologies and guidelines, the state of the art in the field of voice applications and speech interfaces and the evaluation of the applications at different stages throughout development. Investigations of the latest technical development were made as well as an overview of speech interfaces related to automotive on the market. Exploration of user needs via interviews, focus groups and questionnaires was performed. The results from the explorations laid the foundation for the demonstrator work in WP3 where the actual design and development of the prototype for the truck took place. The State of the Art indicated that it was of importance to put more focus on WP2 and WP4 (Vehicle and Safety). Before developing new systems, there is a need to get more insight in how well the voice modality works compared to traditional modalities. This was accomplished in a thorough investigation of user needs thanks to the master thesis students that participated in the project. They had competence within method evaluation and human factors.

### 4.1 State of the Art

The full report containing the State of the Art is found in the report "Automotive Speech Interfaces - A State of the Art survey" by Annika Silvervarg and Arne Jönsson (Appendix 1). It consists of a survey of the latest technical development including recent results from research on design, development and evaluation of dialogue systems in cars, and an overview of speech interfaces related to automotive currently on the market.

Speech interfaces in cars were first introduced in 1996 when Mercedes-Benz S-class car included a dialogue system for operating the car's mobile phone, including number dialing (with connected digit dialog), number storing, user defined telephone directory entry name, name dialing, and directory editing (Heisterkamp, 2001). Since then the development has continued and now all the big car manufacturers have speech interfaces in the cars for tasks like phone, navigation, infotainment and climate control.

Many studies show that the use of mobile phones or physical controls for infotainment systems or climate control are very distracting and can cause crashes so the use of speech interfaces in cars is an important factor for improving safety. Speech interfaces allow the drivers to spend more time keeping their eyes and attention on the road and less time engaging in visual-manually interaction, for example trying to localize the correct buttons and/or navigate through menus. Thus, to allow drivers to interact with voice is seen as an important activity to combat unsafe driving since voice technology aims to reduce the visual-manual load and thus allow the driver to keep his/her eyes on the forward road scene. However, at the same time voice interaction may still impose a relatively high level of cognitive distraction.

The most common applications that are provided with speech interfaces today are phone calls, navigation and entertainment (radio, music). Controlling the cars climate and receiving and sending text messages on the phone are also widely available. With connection to web services it is also increasingly common to be able to find information about Points of Interest (POIs), e.g. businesses like restaurants. There is a big gap between research systems (prototypes) that allow for more natural dialogue and what is commercially available in cars today that are more command based.

## 4.2 User needs

More details regarding the findings of user needs can be found in the reports "Concept Design for Voice Control in Trucks - What do drivers actually want?" by Jonatan Andersson (Appendix 2) and "Speech recognition technology in trucks: potential uses and implications for visual-manual distraction" by Sofia Lindvall (Appendix 3).

Focus group interviews performed within the project with both **developers at Volvo** and **test drivers in Hällered** had the purpose of collecting a first set of speech functionality that were considered useful and could potentially increase the safety of driving trucks. The focus groups were followed up with participant observations. In addition to the interviews, focus groups, and participant observations, 70 phone interviews were held with truck drivers with the ambition to find new possible functions, as well as being able to elaborate the previously found ones. Ten test participants were recruited from each of the seven different truck driver segments in order to identify and prioritize the functions for each one.

## 4.3 How should speech interfaces be designed and implemented?

A big problem with services provided by speech interfaces is that the users might not know the commands or requests that can be made, i.e. what functionality the system provides. They may not know the correct words or phrasing to use or what the correct format for data entry is. This leads to frustration when the user cannot get the information that is available. There are several strategies that can be used to improve on this. A general guideline for speech systems is that the system should match the input vocabulary and grammatical complexity to the output the system gives, as to set the right user expectations. When speech is combined with a graphical interface the words shown on the screen should also be possible to use for spoken interaction. Another way of setting the right user expectations is to match the quality of the system voice to the quality of the speech recognition. A more "robotic" voice can lower user expectations while a very human-like voice will raise them. To help the user learn what can be said and how it should be expressed the system can coach novice users and provide more frequent and more detailed prompts. When the user has become an expert the system should adapt the interaction to allow for more efficient interaction.



There are several things that can be done to increase the naturalness of the interaction and create a more human-like dialogue. One is that the system could always be listening and that the user does not have to press a button and wait for a 'system ready' indication. Another way is that the turn-taking between system and user should allow for barge-in, where the user can skip a system prompt, preferably by speaking over it but otherwise by pressing the push to speak button. Many of today's systems combine speech and touch screen thus allowing for multimodal input and output. In best cases the user can combine modalities freely and use speech together with manual input, for example pointing in a map and saying "I want to go here".

#### **4.4 Voice interaction demonstrator**

The development with the voice interaction demonstrator started with the results from the user needs identified in WP2. The plan for WP3 was to also look into how the voice control interaction can be improved. This was however only done in theory and never tested in the prototype. The work in WP3 resulted in a voice prototype that was implemented into a truck demonstrator used for expert and end user evaluations (performed in WP4).

The dialogue engine was designed with the aim to make the dialogue as natural as possible with the technical boundaries available. One task in WP 3 was also to develop a 'Voice Notification System' to read out certain warning messages which are currently displayed in the instrument cluster as tell-tales as well as other information. In this way the driver would not be visually distracted by looking at the cluster. This feature was not implemented and therefore not tested but was included in one of the concepts on early design phase.

The voice recognition, text-to-speech prototype resulted in a complicated but powerful system which uses Nuance VoCon/Nuance Vocalizer Expressive. It can handle several languages but the prototype is programmed for English and Swedish. It has imperative and dictation modes, where the dictation mode is online which unfortunately is followed by a slight delay in processing. The voice system consists of three programs that handle different aspects of the system. The system is integrated with the truck and uses the cluster and speakers for interaction with user. Communication is started by a Push-to-talk button (in this case the button was placed on right hand side arm rest), the user is then given both an auditory and visual cue to start talking. The user then issues an order and the truck responds (dialogue).

#### **4.3 Methodology and evaluation**

Investigations on how current evaluation methods and tools of voice interfaces can be improved were made and also evaluations of the prototype and its functions from an end-user perspective. The user-related tests methods included both subjective and objective measures and had an iterative process that opened up for improvements in the implementation and evaluation methodology.

For the objective measures, two methods were used: TDRT (Tactile Detection Response Time) and eye tracking. DRT has been widely applied to the measurement of effects of driving and secondary task demand. The tactile version measures the time it takes for a subject to respond to a physical stimulus, in this case a small vibration on the collar bone. For example, it was the main method used in the major Swedish mobile phone investigation (Patten et al., 2003). It is currently subject to standardization in ISO TC22/SC13/WG8 (where the specific task force is co-chaired by Volvo). The ISO standard focuses specifically on the assessment of secondary task demand, in particular related to driver-vehicle interfaces. More on the DRT method can be found in “Using Tactile Detection Response Task for evaluating in-vehicle systems in trucks” by Sofia Lindvall (Appendix 4).

The eye tracking glasses was used to measure eyes-on-road and compare between the different modalities to detect the differences between driving only and driving when interact with voice system or when interact with the traditional systems in the vehicle (voice/visual-manual).

The chosen methods to measure subjective workload were Driver Activity Load Index (DALI), a questionnaire that derives from the NASA TLX form and is design to measure subjective cognitive load when driving. Furthermore, Subjective Usability Scale (SUS) and SASSI (Subjective Assessment of Speech User Interfaces) were used for the subjective evaluations. SASSI addresses following main factors in user’s perception of speech systems:

- System response accuracy (user’s perception of accuracy and function expectations)
- Likeability (usability, friendly)
- Cognitive demand ( perceived amount of effort)
- Annoyance (repetitive, boring, irritating, frustrating)
- Habitability (extent to which users know what to do and what the system is doing)
- Speed (response of user inputs) [Hone & Graham, 2000]

SASSI and SUS were combined to a questionnaire targeting both voice and manual interfaces and took into account visual, manual and cognitive load. The new form was called SUSSI and contained 19 questions. The reason being, that even though SASSI is design with voice interaction in mind, it is not a validated method. SUS on the other hand is validated but not optimized for voice. Both questionnaires had some common or at least similar questions, so these were used with the addition of a few other relevant subjects. More about these subjective forms can be found in “Subjective Usability Measurement for Speech Dialogue Systems In Trucks - A Methodology Study” by Jonatan Andersson (Appendix 5).

The methodology defines independent and dependent measures for efficiency, distraction (both visual and cognitive) and user acceptance.

## 5. Results and deliverables

### 5.1 Delivery to FFI-goals

To design a safe way to interact with the trucks functions as well as with nomadic devices is a serious challenge. Driver inattentions plays a great role in traffic accident on many occasions, however the main focus for research so far has been on visual demand. The RIVER project aimed to facilitate the research about the actual mental workload while using a speech interface and by doing so contributed to the following sub-program goals:

- Analysis, knowledge and enabling technologies
- Basic safety features of vehicles
- Driver support and related interfaces between driver and vehicle as well as interfaces between other road users

WP2 have provided a list of user needs, including how the users are communicating with the truck today, what features in the truck would they like to be able to control with voice, and how can Volvo's voice control interaction be made more efficient to meet these demands. WP2 also investigated how apps and features related to nomadic devices could be safely integrated via the voice control system. WP2 is summarized in Appendix 6.

With WP3, a prototype with a 'natural speech' voice system was implemented in a truck which was evaluated in WP4. A comparison between objective data and data gathered from subjective questionnaires was made in order to evaluate the objective methods.

## 6. Dissemination and publications

### 6.1 Knowledge and results dissemination

The project has contributed to transfer knowledge about the cognitive load during interactions with a voice system.

The project has increased the knowledge about the used methods (TDRT, eye tracking) and this will be beneficial for all participating organisations during future studies.

The project has developed a new form for measuring perceived cognitive work load by combining SAS and SUSSI, and thus creating something more relevant for these types of clinics and studies - SASSI.

The project has compiled a state of the art which can be used in order to develop a better voice control system in Volvo vehicles. Also, a list of user needs has been compiled. By comparing this to existing systems as well as to the state of the art, it gives a finger point towards the direction the commercial vehicle industry should take.

### 6.2 Publications

- Annika Silvervarg, Sofia Lindvall, Jonatan Andersson, Ida Esberg, Christian Jernberg, Filip Frumerie, Arne Jönsson: "Perceived Usability and Cognitive Demand of Secondary Tasks in Spoken Versus Visual-Manual Automotive Interaction"; proceedings of Interspeech, San Francisco, USA, 2016.
- Sofia Lindvall: "Speech versus visual-manual interfaces in trucks: effects on driver distraction, user acceptance, and perceived efficiency"; Master Thesis, Linköping university, Sweden, 2016. (Appendix 7)

## 7. Conclusions and future research

The RIVER project provided a good insight to the current state of art regarding voice interaction in commercial vehicles. The state of art report show how commercial vehicles currently use voice control in their products, this state of art report can easily be updated on a regular basis for a continuous update on voice control in the automotive industry.

The user needs analysis show what the drivers feel they need and want. This result is specifically important for commercial vehicles as voice control in the automotive industry is commonly assumed to be designed for passenger cars. For commercial vehicles the features and surroundings can be drastically different which can totally change the users' needs.

The overall response when using the developed prototypes voice control was generally positive. The perception of the voice system was that it increased safe handling of in-vehicle functions in comparison to the visual-manual type of interaction. Some of the methods for evaluation gave positive results when using the voice control system, however they need to be used in more studies in order to be validated.

The designed voice system was a basic prototype with limited possibilities to interact with. It was designed and implemented for this project and therefore custom made. For example, it only had one person with the same name to call (i.e. "Call Filip" did not require any further actions or possible responses, such as "Did you mean Filip F or Filip S?" or "Did you mean Filip Home or Filip Work phone?"). A larger study to validate results would have to include a voice engine with more data. It would also be beneficial to use, or simulate the usage of, hardware more similar to the one used in commercial vehicles as this might risk lagging responses which would affect the users' opinion.

Another factor that might affect the result was the test track used when performing the evaluations; it was very familiar to the drivers as they had driven it several times before. A study to determine how much the surrounding environment affects the perceived cognitive load of the system could be beneficial.

The drivers only had the choice to do either voice OR manual controlling. There is a possibility that some tasks are performed better with a combination of the two, this is also something that could be tested in future projects.

## 8. Participating parties and contact person



### Partners

#### **Volvo Group Trucks Technology, Advanced Technology and Research (GTT ATR)**

GTT ATR is the research organization within the Volvo Group. Our main assignment is to drive research, development and advanced engineering for the long term competitiveness of the Volvo Group products and services. Our areas of operation also include innovation, business engineering, and corporate services. The group Driver Environment and Human Factors contributes to develop safe and comfortable vehicles and to achieve customer and user satisfaction. The work is based on the customers' and the users' needs and requirements.

Contact person: Filip Frumerie

#### **Linköping University (LiU)**

The division of Human Centered-Design is a highly interdisciplinary division within the Department of Computer and Information Science at Linköping University. Areas of research include human-computer interaction, multimodal interfaces including natural language and augmented reality, interaction and service design, collaborative and social computing, as well as ubiquitous and mobile computing, and accessibility. Much research and teaching is also carried out in Cognitive Science, for instance artificial intelligence including knowledge representation, machine learning, and natural language processing. Finally, cognitive systems, where research includes human factors, cognitive ergonomics, human-machine interaction as well as command and control, emergency and disaster management, safety management, and resilience engineering.

Contact person: Annika Silvervarg

## 9. References

- Apple Carplay (2014). <https://www.apple.com/ios/carplay/>. Accessed 2014-10-14.
- DICO (2009) DICO. VINNOVA P28536-1.  
<http://www.vinnova.se/sv/Resultat/Projekt/Effekta/DICO/>. Accessed 2014-10-14.
- Engström, J. 2011. *Understanding attention selection in driving: From limited capacity to adaptive behavior*. PhD Thesis, Chalmers University, Sweden.
- Frumerie, F. (2013) *SIDVI – Safe and Integrated Driver-Vehicle Interface*. Public report. [http://www.vinnova.se/PageFiles/751290059/2009-03249\\_publicrapport\\_EN.pdf](http://www.vinnova.se/PageFiles/751290059/2009-03249_publicrapport_EN.pdf). Accessed 2014-10-14.
- Herrera, D. (2014) *European Consumer Perceptions toward Infotainment and Telematics - Willingness to Pay for Infotainment and Telematics*. Frost and Sullivan report. MA2D-18.
- Hickman, J.S., Hanowski, R.J., & Bocanegra, J. (2010). *Distraction in commercial trucks and buses: Assessing prevalence and risk in conjunction with crashes and near-crashes*. FMCSA-RRR-10-049
- ISO/CD 17488 *Road vehicles - Transport information and control systems - Detection-Response Task (DRT) for assessing attentional effects of cognitive load in driving*.
- ISO 16673:2007 *Road vehicles - Ergonomic aspects of transport information and control systems - Occlusion method to assess visual demand due to the use of in-vehicle systems*.
- ISO 15007-1:2002 *Road vehicles - Measurement of driver visual behavior with respect to transport information and control systems - Part 1: Definitions and parameters*.
- ISO 15007-2:2001 *Road vehicles - Measurement of driver visual behavior with respect to transport information and control systems - Part 2: Equipment and procedures*.
- Klauer, S. G., Dingus, T. A., Neale, V. L., Sudweeks, J. D., & Ramsey, D. J. (2006). *The impact of driver inattention on near-crash/crash risk: An analysis using the 100-car naturalistic driving study data*. Washington, DC: National Highway Traffic Safety Administration.
- Klauer, S. G., Guo, F., Simons-Morton, B. G., Ouimet, M. C., Lee, S. E., & Dingus, T. A. (2014). *Distracted Driving and Risk of Road Crashes among Novice and Experienced Drivers*. *New England Journal of Medicine*, 370(1), 54-59. doi:10.1056/NEJMsa1204142
- MoU Car 2 Car (2011). *Memorandum of Understanding for OEMs within the CAR 2 CAR Communication Consortium on Deployment Strategy for cooperative ITS in Europe*. Version 4.01.02.
- Olson, R.L., Hanowski, R.J., Hickman, J.S. & Bocanegra J. (2009). *Driver distraction in commercial vehicle operations*. Department of Transportation, Federal Motor Carrier Safety Administration, Report No. FMCSA-RRR-09-042.
- Reimer, B., Mehler, B. (2013). *The Effects of a Production Level “Voice-Command” Interface on Driver Behavior: Summary Findings on Reported Workload, Physiology, Visual Attention, and Driving Performance*. White Paper 2013-18A
- SICS (2014) *Safe Interaction, Connectivity and State*. FFI 2012-04634.



Strayer, D. L., Turrill, J., Coleman, J. R., Ortiz, E. V., Cooper, J. M. (2014). *Measuring Cognitive Distraction in the Automobile II Assessing In-Vehicle Voice-Based Interactive Technologies*. AAA Foundation for Traffic Safety.

Trent Victor, et al. (2014) *Analysis of Naturalistic Driving Study Data: Safer Glances, Driver Inattention, and Crash Risk*. SHRP 2 Safety Project S08A.

Visiongain (2014). *Commercial Vehicle Telematics Market Forecast 2014-2024*.

Volvo vision (2014). *Safety as a core value*.

<http://www.volvogroup.com/group/global/engb/volvo%20group/ourvalues/safety/Pages/safety.aspx>. Accessed 2014-10-14.

US DOT (2014). *Voluntary guidelines reduce visual-manual distraction - the greatest safety risk to drivers in NHTSA's new study*. <http://www.distraction.gov>. Accessed 2014-10-14.



FORDONSSTRATEGISK  
FORSKNING OCH INNOVATION



# Automotive Speech Interfaces – A State of the Art survey

Annika Silvervarg and Arne Jönsson

## Contents

<b>1</b>	<b>Introduction</b>	<b>2</b>
<b>2</b>	<b>Services/Functionality</b>	<b>3</b>
<b>3</b>	<b>Technology</b>	<b>3</b>
3.1	Nuance Dragon Drive . . . . .	4
3.2	CarPlay . . . . .	5
3.3	AndroidAuto . . . . .	6
<b>4</b>	<b>State of the Art systems</b>	<b>6</b>
4.1	In cars today . . . . .	6
4.1.1	Ford SYNC . . . . .	6
4.1.2	Toyota Entune . . . . .	7
4.1.3	GM MyLink . . . . .	7
4.1.4	BMW Connected drive . . . . .	7
4.1.5	Mercedes-Benz . . . . .	8
4.1.6	Chrysler UConnect . . . . .	8
4.1.7	Kia . . . . .	8
4.1.8	Honda . . . . .	9
4.1.9	Volkswagen . . . . .	9
4.2	Research systems/Prototype systems . . . . .	9
4.2.1	CHAT . . . . .	9
4.2.2	Ford Model-U . . . . .	11
4.2.3	SENECA . . . . .	12
4.2.4	VICO . . . . .	13
4.2.5	SAMMIE . . . . .	13
4.2.6	Talkamatic . . . . .	14
4.2.7	CU-Move . . . . .	15
<b>5</b>	<b>User studies and evaluations</b>	<b>15</b>
5.1	Commercial systems and consumer reports . . . . .	15
5.2	Evaluations of research systems . . . . .	16

# 1 Introduction

Speech interfaces in cars were first introduced in 1996 when Mercedes-Benz S-class car included a speech dialogue system for operation of the car's mobile phone, including number dialing (with connected digit dialog), number storing, user defined telephone directory entry name, name dialing, and directory editing (Heisterkamp, 2001). Since then the development has continued and now all the big car manufacturers have speech interfaces in the cars for tasks like phone, navigation, infotainment and climate control.

The use of speech interfaces in cars is an important factor for improving safety. Many studies show that the use of mobile phones or physical controls for infotainments systems or climate control are very distracting and can cause crashes. Ei-Wen Lo and Green (2013) reported that 19% of crashes due to distractions were caused by cell phones, adjustment of infotainment and climate control. Other studies of real-world crashes and near-crashes have consistently demonstrated negative effects of visual distraction, for example, when dialing or texting on a mobile phone.

Studies indicate that these risks can be reduced with speech interfaces. Speech interfaces allow the drivers to spend more time keeping their eyes and attention on the road and less time engaging in visual-manually interaction, for example trying to localise the correct buttons and/or navigate through menus. Ei-Wen Lo and Green (2013) report that it takes 82% less time to input an address using a speech interface compared to traditional keyboard input. They also present several studies which show that lane keeping is better when using speech interfaces compared to visual manual interfaces. There are also studies that show better peripheral detection time and brake reaction time for speech interfaces, as well as lesser subjective workload, and fewer and shorter glances.

But interacting through a speech interface is a secondary task to driving and thus cognitive demanding (Dahlbäck and Jönsson, 2007). A study showed that the reaction time increased by 180ms when using a menu based e-mail system with several options and three levels. Even though speech often is more effective for input to a system, task completion time may be longer for a speech interface since the system can mis-recognise the input and the user may have to repeat himself or correct the system (Ei-Wen Lo and Green, 2013). Problems with automatic speech recognition (ASR) is frustrating for drivers and these problems have increased between 2008 and 2012 (Ei-Wen Lo and Green, 2013). A recent study has shown that the accuracy and reliability of the voice recognition software significantly influences the level of cognitive distraction. Although speech recognition accuracy has been found to affect driving and task performance, no research has been reported on drivers' responses to errors, how long drivers need to take to correct errors, or what strategies drivers use to correct errors (Ei-Wen Lo and Green, 2013).

Thus, to allow drivers to interact with voice has been seen as one important activity to combat unsafe driving since voice technology aims to reduce the visual-manual load and thus allow the driver to keep his/her eyes on the forward road scene. However, at the same time voice interaction may still impose a relatively high level of cognitive distraction. The purpose of this report is to survey what has and can be accomplished with speech interfaces in cars. It will focus on the latest technical development and provide an overview of speech interfaces related to automotive on the market, but also include results from research on design, development and evaluation of dialogue systems in cars.

## 2 Services/Functionality

What kind of applications are provided with speech interfaces in today's cars? The most common are phone calls, navigation and entertainment (radio, music). Controlling the car's climate and receiving and sending text messages on the phone are also widely available. With connection to web services, finding information on Point of Interest (POI), e.g. businesses like restaurants, is also an increasingly common service. These are illustrated by example dialogs below that are representative of what can be found in cars on the market today:

S: Command please.  
U: Find address.  
S: Find address. The city please. (Default State is Michigan)  
U: Michigan City.  
S: Please select the respective line or start spelling.  
U: M I C H I G A N.  
S: Spell the name in blocks of letters or enter your destination again with change state or enter city.  
U: Change state.  
S: Please enter the state name.  
U: Indiana.  
S: Please select the respective line or start spelling.  
U: Line one.  
S: Line one. Indiana. The city please.  
U: Michigan City.  
S: Michigan City. What is the street?  
U: Wabash Street.  
S: Wabash Street. And house number or if you do not know that please say show map or start guidance.  
U: Six zero one.  
S: Six zero one. Show map or start guidance.  
U: Start guidance  
S: Start guidance.

S: Please say a command.  
U: USB.  
S: USB. Is it correct?  
U: Yes.  
S: USB. Please say a command.  
U: PLAY TRACK "Like a Prayer."  
S: Play track "Like a prayer."

S: Main menu.  
U: Phone.  
S: Phone. Please say a command.  
U: Call XXX XXX XXXX.  
S: Calling XXX XXX XXXX.

## 3 Technology

A survey of how speech interfaces operate in cars on the market, shows a very conform design over all brands. Typically the cars have a push to talk function with a button on the steering wheel (in some cases there are two different buttons where one is dedicated to access the phone's functionality). A system ready mode is conveyed through a beep that signals that the user can talk. As illustrated by the examples in the previous section the driver can then enter a command (or the system prompts

the user, for example "Please say a command"). Often the initial command is one of navigation, phone, music, radio, climate. After the user initiates the system takes over and prompts the user for further information. The dialogue is very simple and could be modelled by a finite state machine where the system step by step collects attributes to fill a form/frame corresponding to a specific service/command.

The user is restricted in the way information can be input, for example how addresses should be expressed or how number sequences in telephone numbers or radio stations can be expressed, see examples above. However, there is a development towards more free speech which allows the user to both state all or many pieces of information in one utterance and to use more natural phrasings of requests.

Many of the systems are multi-modal with an integration of voice and touch screen. That means that the user can choose to input information also by touch, as seen in the example with navigation where the user is offered to choose a line (from a list on the display). Some systems offer full multi-modality, i.e. the user can always choose between the two modalities for input, while some restrict the user to use only one modality in some cases, for example to choose one alternative from a list.

The following sections will describe the most commonly used technical platform which is Nuance Dragon Drive and the new trend of integration of mobile phones and their voice interfaces to various applications, similar to those provided in cars, i.e. entertainment, navigation and phone.

### 3.1 Nuance Dragon Drive

Nuance is one of the biggest companies and providers of speech and language technology and Nuance Automotive has since 1996, when they were the first company to provide speech input in a car, also provided the first voice destination entry in 2005 and first in-car message dictation in 2012. In 2012 Nuance were shipping in more than 20 million cars in brands such as Audi, BMW, Chrysler, Fiat, Ford, GM, Hyundai, Mercedes, Toyota, Volkswagen.

Dragon Drive (Dragon, 2013) is Nuance's platform for speech interfaces in cars. Dragon Drive can be used for services like navigation, local business search, music, telephone and messages, climate control and more. Dragon Drive consists of three parts: Dragon Drive Speech, Dragon Drive Connect (Schmand, 2014) and Dragon Drive Link.

Dragon Drive Speech utilises Nuance VoCon 4.5 and Vocalizer Expressive ASR and TTS. Their latest version of the VoCon speech recogniser allows for much improved interaction, which has been announced by auto makers to be introduced in the near future i.e. 2015/2016 year models. For example, it recognizes natural speech, eliminating restriction to predefined commands and enables all commands to be spoken in a single utterance on the main menu. This is done using semantic classification of the recognised words. It can also handle recognition of information from combined large lists recognizing only valid combinations, e.g. street + city + state for all USA. It recognizes every possible word in a database in every possible order and permutations, including partial utterances. This search is especially useful when the user does not know the exact wording of the content he or she is searching for.

The new VoCon version also eliminates the push to talk button as it has an always listening mode that allows the user to wake up or activate their system with a keyword, like Hello Dragon. Another feature is the handling of barge-in, i.e. that users interrupt and talk over system prompts. This allows a user to speak at any time during a dialogue. Imagine someone is sending a text message by voice. After finishing dictating, the system might read back a final confirmation - Please confirm: your text will go to John Smith. It says: Hi John, are you available... Rather than

waiting until the whole text message is read, or having to hit a steering-wheel button to confirm by voice, the user can simply speak, "OK send it", at any time.

Dragon Drive Connect allows for connection to third-party applications and content, such as Yelp, Rdio, AccuWeather, OPIS/ Gas Buddy, INRIX, TomTom, MovieTickets.com, Reuters, Twitter, Facebook, available over Dragon Drive Link that provides the link between the car and a mobile phone (Schmand, 2014). It connects the car with the cloud where real-time content retrieval is done. A cloud-based Transform Engine is utilised to process users voice commands into information requests and decide on appropriate actions through dialogue, such as clarification requests for ambiguous requests, or missing pieces of information to complete a vague request. Thus, both dialogue management and domain reasoning is done in the cloud and can be maintained and updated off-board the car.

Nuance has also announced the Dragon Drive Daily Update, a virtual assistant service providing personalized content to the driver. Using voice biometrics a spoken pass-phrase can be used to identify the driver provide content and information, including connected music stations, navigation routes, traffic conditions, calendars, and more, based on that driver's preferences. Nuance is also expanding its Dragon Drive platform with the Dragon Drive Mobile application, supporting iOS and Android, which let automakers build a mobile app that integrates directly with their customized and branded in-car infotainment systems. The app delivers the same app experience on the phone and in the car for navigation, music, calendar, dictation and more. In addition, drivers can set their music preferences, navigation, and other information using the app on their phones, and upon entering the vehicle, the same music and navigation information is accessible via voice. Once connected, the application locks the driver's smartphone and transfers all interactions to the in-car head unit screen.

## 3.2 CarPlay

One of the best-known non automotive natural language speech interfaces is Siri, released by Apple in October 2011. With the assistance of Siri users can make phone calls, find a business, get directions, search the web, and perform other tasks supported by apps on iPhones. In 2014 Apple announced the CarPlay system that features Siri voice control and is specially designed for driving scenarios. The aim of CarPlay is to provide direct access to iOS device functionality through the car's controls - knobs, buttons, or touch screen. The apps used in the car can be redesigned, so they do not require the user to look at the screen. Siri enters a driving mode when connected. This means not every task you can do with Siri is available (like making reservations with OpenTable or looking up Wikipedia articles), but other tasks understand immediately that you can't see your iPhone display and persistently read back text to you (like setting reminders and sending emails).

Integration is aimed for several functions that iOS devices currently incorporate, these include: Siri: Eyes Free mode - for eyes-free and hands-free operation Satellite navigation (Satnav), Telephony instruction and control, Music control (via Apple's iOS "Music" application, or third-party), iMessage control and response. The initial time scale for release is from 2014. The first who were on-board were Ferrari, Mercedes and Volvo, but others with announced plans include BMW Group, Ford, General Motors, Honda, Hyundai Motor Company, Jaguar Land Rover, Kia Motors, Mitsubishi Motors, Nissan Motor Company, PSA Peugeot Citroen, Subaru, Suzuki and Toyota Motor Corp. Since most have multiple brands, it's more than two dozen car brands. The biggest missing name is Volkswagen Group, the world's number three car manufacturer, which also includes Audi, Bentley and Porsche.

### 3.3 AndroidAuto

Android Auto is part of the Open Automotive Alliance, which was announced by Google, GM, Honda, Audi, Hyundai, and Nvidia, on January 6, 2014, and is a joint effort with 28 automobile manufacturers and mobile tech supplier Nvidia. The Android Auto mobile app was released on March 19, 2015. The aim of Android Auto is to extend the functionality of an Android mobile device in an automobile to the dashboard's head unit. In order to use the system, users must be running Lollipop on their mobile device and must own a vehicle supporting Android Auto. Several functions are supported: GPS mapping/navigation, Music control, Telephony, SMS composition and playback, Web search. Most of the features will be controlled through Google Voice, with voice command processing done by your phone. Instead of an icon-based interface, as CarPlay uses, Android Auto relies on panels of what it considers relevant information. There are icons at the bottom of the screen, however, working as shortcuts to navigation, phone, and audio.

An Android Auto software development kit (SDK) will be released to developers soon, allowing third parties to modify their apps to work with Android Auto. APIs will initially only be available for music and messaging apps.

## 4 State of the Art systems

There is a big gap between research (prototype) systems that allow for more natural dialogue and what is commercially available in cars today that is more command based. Therefore the State-of-Art section has two parts, one on commercially available systems and one on research systems.

### 4.1 In cars today

#### 4.1.1 Ford SYNC

Ford SYNC was initially built on Windows Embedded Automotive platform (Ghangurde, 2010), but since 2007 it uses Nuance for speech recognition and text to speech. SYNC was one of the industry's first systems to widely and affordably offer speech interfaces in cars. Using commands in multiple languages, such as English, French or Spanish, drivers can operate navigation, portable digital music players, and Bluetooth-enabled mobile phones. Many of the examples in Section 2 are authentic interactions with SYNC. They typically consist of the user initiating a command, e.g. "Climate", the system echoing the command (to indirectly confirm it has correctly understood the user), followed by a more specific command, e.g. "fan", the system echoing a confirmation or prompting for more information, e.g. "fan temperature please", and the user providing the desired information, e.g. "maximum".

Ford has announced SYNC3 to be introduced in their 2016 and 2017 year models. It is to have a more natural interaction by minimizing the number of steps needed to carry out a command. The user can even be a bit vague and still be understood saying "play song/ artist/album/genre" will have the system play the desired song, artist, playlist or album, the user will not even have to identify the desired category. The increased capability in the system's natural language understanding means that the users do not have to know an exact name. They can search for "Detroit Airport" rather than using the official name "Detroit Metropolitan Airport." With addresses, they can say, "Eleven Twenty-Five Main Street" instead of "One One Two Five Main Street."

SYNC 3 also allows for better integration with smartphones. With AppLink customers can connect their smartphone to their vehicle and control their compatible apps using voice commands or buttons on the vehicle display screen. AppLink now automatically discovers smartphone apps including Spotify, Pandora, Stitcher, NPR One, SiriusXM Radio and iHeartRadio Auto, and displays their unique graphics and branding. Music and news apps are automatically displayed along with other media sources, just like AM/FM or SiriusXM. Ford will support Android Auto and CarPlay eventually.

#### **4.1.2 Toyota Entune**

Toyota Entune (Toyota, 2015) premiered in 2011 on the 2012 model year Toyota Prius V and is available in selected Toyota cars. A mobile phone running the Entune App Suite can connect to the car and the apps can then be operated through the cars controls, or for some by voice recognition. The apps include information on traffic, weather, sports scores, stocks, and fuel prices via subscription through SiriusXM.

Before using the voice recognition the user is recommended to train the system and run a tutorial. To use the speech interface for dialling, destination input, or searching a point of interest, the use needs to press the push to speak button and answer the system's prompts. For destination input it is also possible to state the entire address in one action. Numbers must be entered one digit at a time. If a search results is a list of alternatives it is displayed on the screen and the user selects by saying the number for the chosen item. A prompt can be interrupted by pressing the push to speak button, thus allowing for barge in.

Toyota has also announced a new "Advanced Voice Recognition" which will bring a number of new features. All applications can be launched using voice commands and all applications except Facebook, Places, and Fuel Guide can be operated using Advanced Voice Recognition. Over 100,000 variations of commands are recognized, and the user will be able to speak in a free form. There is also a Hint Screen on the vehicle navigation system that displays sample commands that can be spoken. The user will be prompted for smaller pieces of information to clarify a request if needed.

#### **4.1.3 GM MyLink**

MyLink aka Intellilink is a telematics system/infotainment system offered by General Motors vehicles starting from the 2012 model year. It uses voice recognition software from Nuance to control online services, like Pandora Internet Radio or Stitcher Radio through voice commands. Speech interfaces are also provided for navigation and phone. The new version does not rely on a fixed set of commands but can instead interpret full sentences and common phrases. "Play the radio," will elicit a response asking the user to specify AM or FM or to simply choose a station. Asking the car to change the destination results in the system asking the user if she wants to use an address or look for a nearby point of interest.

Chevrolet is the first vehicle manufacturer to offer Siri Eyes Free to compatible iPhone users on their 2015 model year.

#### **4.1.4 BMW Connected drive**

In BMW cars voice can be used to control features such as the telephone, climate control, navigation and sound systems with spoken commands. A contact from the telephone directory can be selected by saying the name and a telephone number can be dialed by just saying the number. The text-to-speech function can read aloud emails or text messages. A dictation function with speech recognition allows the

driver to dictate short emails and text messages by Voice Control. Voice memos up to two minutes long can be recorded and sent by email. The system recognises preset terms with commands for nearly all functions and always follows a predefined dialogue structure. But it has some flexibility as it allows for 'one shot' complete entry where instead of entering the destination in single steps as town, street and number, the whole address can be said in one sentence, e.g. "Munich Riesenfeldstrasse 7". The system is also able to recognise alternative wordings, e.g. "Home" instead of "Drive home". The voice control system is activated by pressing a key on the steering wheel and the system emits a tone to indicate it has been understood.

Voice can also be used to 'navigate' through the menus on the central display. Say the name of the menu to go there, e.g. "Main menu" or "Radio". The command "Map" automatically opens up a map in the Control Display. Most menu options in a displayed table can also be called up by voice control. For example, you can change the map's orientation with a spoken command.

#### **4.1.5 Mercedes-Benz**

Daimler-Benz, later Daimler-Chrysler, have participated in many research projects and were the first to introduce a speech interface in a car in their Mercedes-Benz S-class already in 1996 in Germany and 1999 in the US (Heisterkamp, 2001). The first version could be used for phone and had a vocabulary of 30 words. It was later extended with climate control and a vocabulary of 300 words. They used an in-house developed HMM-based speech recognizer with high accuracy. But since 2007 they are also using Nuance' speech recogniser.

Voice control is activated by pressing the Push-to-talk button on the steering wheel. It enables operation of the most important functions - navigation, music and telephony - by voice input. The optional COMAND Online\* system offers internet access via a mobile phone connection while stationary, as well as free use of Mercedes-Benz services such as weather, POI search and route download via Google whilst driving. The highlight of the latest generation of LINGUATRONIC is so-called whole-word voice input. This means that the driver is no longer required to spell out his wishes, instead stating what he wants in simple terms when it comes to specifying a destination, choosing a radio station or calling up a name from the stored phone book.

#### **4.1.6 Chrysler UConnect**

Chrysler UConnect (UConnect, 2015) supports phone, navigations, media, radio, climate and other apps available through Travellink or UConnect Access. It uses push to talk with two different buttons, one of which is dedicated to the "Tune to ....", "Change source to iPod", and the system then prompts for more specific information. The default service for text messages is realised by text to speech for reading aloud of incoming messages and the option to answer the message with one of 18 predefined answers. If the user pays for extra apps voice texting is included.

#### **4.1.7 Kia**

Kia utilizes a complete Microsoft solution, integrating the Tellme speech recognition engine. Windows Embedded Automotive is an operating system subfamily of Windows Embedded based on Windows CE for use in computer systems in automobiles. The platform focuses on infotainment and gives voice-directed control over music and phone. Users can give voice commands without having to navigate through menus. By supporting complex grammar, UVO needs only short voice commands to connect



drivers and passengers with their desired functions. To improve upon the system's default capabilities, the user is given the option to create voice profiles as a training tool for Microsoft's engine. The profile creation process requires that the user reads a couple of sentences, allowing the software to detect the user's speech patterns and voice inflection.

#### 4.1.8 Honda

Honda is moving towards free speech, the request "Radio 98.3 FM" can now be spoken in approximately a dozen different ways ranging from "Change the radio station to 98.3" to "98.3 FM." The navigation system can also be controlled by voice with conventional navigation commands like, "Find the nearest Chinese restaurant" or "Find the nearest ATM." A large point-of-interest (POI) database includes telephone numbers that can be dialled by using the Bluetooth HandsFreeLink system when the driver's cellular telephone is connected to the system. A standard SMS Text Message function can read incoming texts aloud, and the driver can reply with any of six factory preset messages ("Talk to you later", "I'm driving", "I'm on my way", "I'm running late", "OK", "Yes", "No"). Honda has also integrated Apple's Siri Eyes Free. Users with compatible iPhones will be able to operate Siri through familiar voice commands by pressing and holding the TALK button on the steering wheel when their iPhone is paired via Bluetooth.

#### 4.1.9 Volkswagen

Volkswagen and The Electronics Research Laboratory (ERL) (which is a part of the global research and development network that supports the Volkswagen Group brands), have participated in the development of the research system CHAT (Chang et al., 2009), see Section 4.2.1. Unfortunately, it is hard to find information regarding the speech technology available in Volkswagen cars and how much of the research results have been implemented. 2010 they announce the collaboration with SVOX on making in-car voice commands even more natural and intuitive. Instead of being strictly guided through a menu, drivers should be able to use their natural, everyday speech. This endeavour was based on the combination of statistical language models and the slot filling dialogue manager from SVOX. The system filters the relevant information from an utterance, regardless of the speakers' wording or word order. A frame-based dialogue system supports the user in effectively achieving the intended action, asking questions where information is missing and providing additional information where necessary. Users will be able to freely correct the system by over-answering and even changing their mind during a command. In 2011 SVOX was acquired by Nuance.

Elektrobit helped develop the navigation and user-experience features in the Volkswagen Golf VII standard infotainment system in 2012, which includes robust voice capabilities. The voice control has been developed with the speech dialogue software EB GUIDE Speech Target Framework and features speech destination entries for more than 50 countries. The voice control for navigation and entertainment support both step by step specification of commands and 'one-shot full word input'.

## 4.2 Research systems/Prototype systems

### 4.2.1 CHAT

The CHAT system (Weng et al., 2006, 2007; Cavedon et al., 2007; Yan et al., 2007) was a joint effort of Bosh RTC, CSLI of Stanford University and ERL of VW of

Figure 1: A dialogue illustrating the capabilities of the CHAT system.

U: Switch to restaurant.  
S: Ok. Switched to the restaurant finder.  
U: Can you find me a Chinese restaurant in Palo Alto with good service?  
S: There are 11 Chinese restaurants in Palo Alto that have good service. Here are the first few: Little Garden Restaurant China Delight Mandarin Gourmet Restaurant Should I continue?  
U: Hmm...tell me more about, Garden Restaurant?  
S: Little Garden Restaurant is an expensive Chinese restaurant. Dress code is casual, service is good and rating is three stars. It's on El Camino Real in Palo Alto.  
U: Switch to navigation.  
S: Ok, switched to Navigation system.  
U: Take me to Little Garden Restaurant, but stop by Bosch Page Mill Office.  
S: Ok, I will plan a route to Little Garden Restaurant that goes by Bosch Page Mill Office. Would you like anything else?  
U: Get me the fastest route.  
S: Ok, I will plan a fast route to Little Garden Restaurant that goes by Bosch Page Mill Office. Would you like anything else?  
U: Um...nope.  
S: Ok, I will plan a route now. Please wait ... I calculated a route to Little Garden Restaurant that goes by Bosch Page Mill Office. It is 4.3 miles long and should take around 12 minutes. Now I am starting guidance...  
U: Switch to music player.  
S: Ok. Switched to mp3 music player.  
U: What songs do you have by Alicia Keys?  
S: There are 16 songs by Alicia Keys. Here are the first few: GoodBye, Jane Doe, and Caged Bird  
U: Play "Caged Bird."  
S: Playing the song "Caged Bird."  
U: Add it to a playlist.  
S: Which playlist do you mean, car or work?  
U: What's on the car playlist?  
S: Interlude, A Womans Worth, and Whatever.  
U: Add it to the car playlist then.  
S: I added the song "Caged Bird" to car.  
U: Switch to navigation.  
S: Ok. Switched to navigation system.  
U: Where am I right now?  
S: You are on Page Mill Road in Palo Alto.  
U: How far away am I from my destination?  
S: You need to travel 2.5 miles to get to destination Little Garden.

America. The system is specifically designed to handle imperfect speech and imperfect memory of the user when using a dialogue system as secondary task while performing a primary task such as driving. This can take the form of speech disfluences, incomplete references to proper names, and phrase fragments. The system provides services for phone, restaurant finding and navigation. The dialogue in Figure 1 illustrates the capabilities of the CHAT system.

The speech understanding includes a partial name identifier, a shallow semantic parser and a deep structural parser. It delivers a list of the n-best alternatives to the dialogue manager, so that it can decide which interpretation is most likely based on the possible dialogue moves and the dialogue context. To deal with possible misunderstandings the system asks for clarification, confirmation or rejection when it is not confident about the interpretation. It may also seek confirmation implicitly and give the user the opportunity of revision by conveying the chosen interpretation in the next step of the dialogue. The dialogue manager has its roots in the CSLI dialogue manager and is based on the information state update approach. It models the dialogue history in a dialogue tree and can thus keep track of multiple dialogue threads. It allows for mixed initiative dialogue in all its domains. Domain knowledge

is modelled using OWL ontologies and are used by a knowledge manager to access the knowledge bases. A content optimizer acts as a mediator between the dialogue manager and knowledge manager as well as the application manager. If the results of a user query are too complex, with too many alternatives, the system proposes additional criteria that help the user narrow down the number of results. On the other hand if no information can be found regarding a request the system proposes relaxation of the constraints. To facilitate the dialogue and support the user, the system can also repeat information previously provided, it can handle anaphora, partial names and ordinal references, such as, the last one or the second one.

#### 4.2.2 Ford Model-U

In 2003 a prototype of an automotive multi-modal dialogue system developed by Ford and Speechworks was introduced (Piebaccini et al., 2004). It aimed at improving the current state of command driven speech interaction by combining speech functionality and a touch screen. Thus the driver should be able to adapt the style of interaction depending on the driving situation and task. The provided services/functionality were navigation (destination entry, point of interest, scroll and zoom map, find current location), entertainment (play MP3s by category, artist name, playlist name, brows list, change volume), telephone (dial by number and name, browse contact lists) and climate control (cabin and seat temperature, fan speed, fan direction, recirculation, fresh air, open/close roof, front and back defrost).

The system uses mixed initiative dialogue. The user can initiate a dialogue by uttering a command, such as "climate control". The system then takes the initiative and prompts the user for more information, e.g. temperature, position. The different alternatives are also shown on the touch screen and the user can choose to answer by speech or touch. All words visible on the GUI can also be spoken. The user can take the initiative during a dialogue with the car and change the mode, for example from climate control to navigation. This can be done by giving a speech command, e.g. "navigation", or through input by the GUI. Thus for inexperienced users the system guides the users through several steps in a directed dialogue. An expert user can provide all the desired information in one utterance instead, for example "climate control driver's seat temperature down". These are effective terse commands. A middle way is to use natural language commands such as "Turn the seat temperature all the way down". If the system misunderstands or information is missing from a command or utterance it will initiate a directed dialogue to get the missing pieces. (Piebaccini et al., 2003)

These examples of user utterances are one type of input signals that can initiate a dialogue. Another is signals from the car itself, for example low-fuel levels. This can initiate a dialogue concerning navigation that guides the driver to a gas station (Piebaccini et al., 2004).

For speech recognition the system uses SpeechWorks' Speech2Go with semantic models of the context and conditional confirmation to improve the accuracy and overall interaction. The speech recognition output strings are scored based on their semantic content and how well they match the current context, so that the most likely semantics get a higher score. The best results are evaluated against thresholds to decide if they should be rejected, accepted or confirmed. The confirmation threshold is adapted depending on the match between the utterance's and context's semantics. If they have the same semantics the threshold is lowered since the utterance is more likely to be true than if it is out of context. (Piebaccini et al., 2003). The ASR accuracy was 93.7% without dynamic semantic modelling and 94.7% with.

The system is based on a multi-modal dialogue manager ETUDE that takes input

signals, i.e. user speech actions in the GUI and/or signals from the car and responds with output actions. The signal is processed by a recursive transition network and updates a frame that holds all the session state variables. The network consists of several sub networks for different kinds of commands and (sub-)dialogues that correspond to filling a frame with all information pieces needed to execute the command. Global transitions allow the system to change the current node, and thus to shift to a different command (Pieraccini et al., 2003).

Ford Model-U was the predecessor to Ford SYNC but it seems that many of its advanced features has yet to be incorporated in a commercial solution.

### 4.2.3 SENECA

The SENECA spoken dialogue system (Minker et al., 2004, 2003) was developed as part of an EU-project where DaimlerChrysler were a part and Temic supplied the speech recognition technology. It is a command-based speech control of entertainment (radio and CD), navigation, and phone, and allows about 75 different commands. Synonyms may be used to form alternatives of a command, such as switch on radio, turn on radio or activate radio. The system uses a grammar in Backus Naur Form to perform semantic interpretations of user utterances resulting in feature-value pairs. These feature-value pair are then used by the dialogue manager. For dialogue processing the system uses a menu-based dialogue strategy, which included both top-down access for main functions and side access for sub functions. The dialogue differs for novices and experts, where novices receive more polite and detailed prompts, while experts get more brief and clear prompts. The mode can be chosen by the user or the system automatically based on the amount of recognition errors.

To handle the problem of how to correctly identify destinations from a database that contains around 69000 city names, that can easily be mistaken since many of them a phonetically confusable, the system relies on clarification dialogues when requests are ambiguous (have many matches) or have low confidence. The first step is to ask the user for a zip code. If this is provided the initial input is re-recognised using a modified vocabulary narrowed down to the cities within the provided zipcode area. If the user cannot provide a zipcode or the original utterance still cannot be uniquely identified as a city, the following steps are to ask for more/other information such as a nearby city or a region, trying to re-recognise the input given the new information. As a last resort the user is asked to spell the city name. A resulting dialogue is presented in Figure 2.

Figure 2: A dialogue illustrating the capabilities of the SENECA system.

S: Yes please?  
U: Enter address.  
S: Name of the city?  
U: Siegburg.  
S: Do you know the postal code?  
U: No .  
S: Do you know a city near the destination?  
U: No.  
S: Do you know the federal state?  
U: Yes.  
S: Name of the Federal state or region?  
U: North Rhine - Westphalia.  
S: <Displays city name> Is the name of the city correct?  
U: Yes.  
S: Navigation to destination started.

#### 4.2.4 VICO

The VICO system (Coletti et al., 2003; Geutner et al., 2002) was developed during an EU project which also involved Daimler-Chrysler. It is a very complex spoken dialogue system that uses sophisticated methods for speech recognition, natural language understanding and dialogue management (including task management and domain management). Dialogues are spontaneous mixed-initiative dialogues initiated by the user, in which the user has the initial initiative and always has the option of taking the initiative, changing the topic or the task at will. Natural language understanding receives sequences of word hypothesis graphs from the speech recogniser and analyses them linguistically using robust ("island") parsing. The parser result is inserted into a semantic frame and passed on to the dialogue manager for further processing. To facilitate the interpretation it utilises expectations as to the next user input. When the system has asked for route input, the user response is parsed by the route sub-grammar and there a number can only trigger a street number slot. Distinguishing between task-dependent and task-independent rules makes it possible to organise the grammar into different parts, which helps limiting the amount of work necessary to write a new grammar for a new task, because only new rules that are relevant to this task are needed.

The dialogue manager receives a semantic frame from the NLU and produces a high level semantic specification of appropriate output to the user. Dialogue management is separated into two main parts, task management and domain handling. Task management includes receiving semantic (user) input frames from the NLU, querying the back-end through the domain manager, sending semantic (system) output frames to the response generator, sending predictions (or expectations) on the next user input to the speech recogniser and the natural language understanding module, and to keep track of the dialogue structures for each task supported by VICO and two different dialogue histories; the topic history and the task history (Bernsen et al., 2002).

#### 4.2.5 SAMMIE

SAMMIE was the result of the EU-project TALK in which BMW and Bosh participated (Becker et al., 2006a,b). It focused primarily on a music application. It is a natural mixed initiative multi-modal dialogue system where the user have control of the interaction and can give input through any modality and do not have to answer system queries, but can initiate new tasks whenever she wants. This is achieved through the use of recipes that acts as models used for collaborative problem solving and AI planning, and enriched information states that capture the dialogue context and the users attention state. This allows for a free format for user input, which can be multi-modal (combining speech and gesture "Play this title"), see example dialogue in Figure 3.

Figure 3: A dialogue illustrating the capabilities of the SAMMIE system.

U: Show me the Beatles albums.  
S: I have these four Beatles albums. [shows a list of album names]  
U: Which songs are on this one? [selects the Red Album]  
S: The Red Album contains these songs [shows a list of the songs]  
U: Play the third one.  
S: [song "From Me To You" plays]

The speech and gesture recognisers provide alternative semantic interpretations that are fused and further interpreted in the context of the dialouge state. The dialogue manager access the music application and decides on an appropriate response

given the results and the current dialogue context and the task (seen as a collaborative problem solving task). The problem solving activities are modelled by feature structures in a hierarchy with 6 abstract problem solving objects that model several domain-independent types that correspond to tasks such as play-song. The collaborative problem solving activity consists of filling the appropriate slots in the structure. An OWL ontology is used for domain modelling of the player, database and playlists, as well as the tasks users can perform. This model is user centred and reflects the users perception of the systems abilities. This can then be mapped on the underlying application.

#### 4.2.6 Talkamatic

The Talkamatic dialogue system provides telephone and navigation services and is based on the dialogue manager GoDiS, developed at the University of Gothenburg (Kronlid et al., 2011). It represents a novel dialogue model, compared to the dialogue systems usually offered in a car setting. The dialogue system allows the user to give little or much information in one utterance and have a rich model for establishing shared understanding (a.k.a. grounding). The dialogue manager is based on the Information State approach and it handles task switching and grounding .

The system can convert menus into dialogue plans. While it is possible for the novice user to let the system take initiative and guide the user through the menus step-by-step, expert users might prefer to use accommodation strategies enabling the user to skip through the menus and getting right to the point. For example when the systems prompts the user "What do you want to do? Go to the phonebook, manage settings, or..." the user can interrupt and say "Call Lisa's home number" without going through any intermediate steps in a system-driven dialogue.

The system also handled arbitrarily nested subdialogues. The user can initiate a new task and then go back to a previous task. See, te example in Figure 4.

Figure 4: A dialogue illustrating the task switching capabilities of the Talkamatic system.

U: Change Lisa's home number.  
 S: Okay. Let's see. What phone number do you want instead?  
 U: Check my messages.  
 S: You have got this message. Hi! I have got a new home number, it is (031)234567. Best regards, Lisa.  
 S: Returning to change an entry. What phone number do you want instead?  
 U: oh three one twentythree fourtyfive sixtyseven.  
 U: Okay. Changing Lisa's home number to oh three one two three four five six seven.

The dialogue manager also provides feedback strategies to make sure that the system have correctly understood the words that are spoken (semantic understanding) and the meaning of the utterance (pragmatic understanding) and that the user accepts the dialogue moves performed in utterances. As an example, the single user utterance "Lisa" may result in positive feedback on the semantic level but negative on the pragmatic, resulting in a system utterance consisting of two feedback moves and a clarification question: "Lisa. I don't quite understand. Do you want to make a call, change an entry in the phonebook, or delete an entry from the phonebook?"

The system implements full mulit-modality where the user can freely switch between and combine modalities across and within utterances. This makes it possible to use the system using speech only, using traditional GUI interaction only, or using a combination of the two. User input can use several input modalities, e.g. "Call this contact *CLICK*" where the *CLICK* is a mouse click. For output parallel multi-modality is used, i.e. output is generally rendered both as speech and as GUI output.

#### 4.2.7 CU-Move

CU-Move was a research system built on the MIT Galaxy-II Hub architecture by the Center for Spoken Language Research (CSLR) at the University of Colorado, Boulder (USA) (Hansen et al., 2005, 2000; Pellom et al., 2001). Their primary goal was to develop algorithms and technology for robust access to information via spoken dialogue systems in mobile, hands-free environments. It mainly focused on multi-channel noise suppression, automatic environment characterization, and on a prototype navigation dialogue.

## 5 User studies and evaluations

As stated in the introduction, speech interfaces in cars have the potential to increase safety through lessening the distraction. But how well the technology works is an important factor in achieving this positive effect. In this section we will report on how well speech interfaces in commercial systems work, and what evaluations of research system have shown as potential challenges that have to be met in design and implementation of good speech interfaces in cars.

### 5.1 Commercial systems and consumer reports

User studies of how speech interfaces in cars are received show that although voice recognition is widely used it does not meet the customers expectations. Almost one-in-four U.S. motorists use voice recognition in their cars daily and 53% tap it at least once a week. But audio, communication, entertainment and navigation (ACEN) systems are reported as the most problematic component category in today's new vehicles. The 2014 Multimedia Quality and Satisfaction Study conducted by J.D Power is based on responses from 86,118 new-vehicle owners surveyed between February 2014 and May 2014. The study measured the experiences and opinions of vehicle owners regarding the quality, design and features of their ACEN systems in the first 90 days of ownership. Multimedia system quality is determined by the number of problems experienced per 100 vehicles (PP100), with a lower score reflecting higher quality.

In recent years, problems with ACEN have become the most prevalent type of problem with new vehicles. Specifically, built-in voice recognition surpasses wind noise as the problem most frequently reported by new-vehicle owners and has risen to 8.3 PP100 in 2014, up from 7.6 PP100 in 2013. The voice recognition problems customers cite most often relate to three built-in hands-free communication issues: does not recognize/misinterprets verbal commands (63%); does not recognize/misinterprets names/words (44%); and does not recognize/misinterprets numbers (31%).

Although a majority of new-vehicle owners continue to express interest in having built-in voice recognition and connectivity, these same owners indicate their wireless phone is more robust than current built-in systems, and they are not eager to pay for technology they perceive will not work as needed or expected. Nearly three-fourths (70%) of new-vehicle owners indicate interest in built-in voice recognition. When given a cost of \$500 for this technology, purchase interest drops to 44%.

Auto manufacturers produce built-in voice recognition and connectivity systems that are not in sync with consumer expectations. It takes three-to-five years to take a vehicle from drawing board to dealership and just a year or so for a handset. "With a car, I'm always way behind the curve, Siri is never behind the curve."

## 5.2 Evaluations of research systems

An evaluation of the CHAT research system (Chang et al., 2009) identified a number of (potentially) problematic areas. They had 30 persons of varying age and experience as drivers conduct 21 tasks using the speech interface in a stationary car. There were 7 tasks on using the telephone, 8 on navigation and 6 on maps. There was a time limit of maximum 3 mins to complete a task before they were interrupted. The measures used were task completion (complete/not complete), task time (seconds), task difficulty (subjective rating from 1 very easy to 10 very difficult). When analysing the tasks that the users had trouble completing, a number of consistent problem areas that affected the interaction were found. These can be formulated as design issues that need to be considered when designing and implementing a speech interface in cars. Some of the issues can be avoided with better speech technology but many of them are still highly relevant based on the available technology today.

**System organisation** The user should have a clear overview of the system's organisation. A complicated structure creates confusion and hinders performance of the tasks.

**Global vs local commands** It should be clear which commands are context dependent and which are not (if there is a this separation in the system).

**Undo or Back** It should be clear how the user can cancel an ongoing task or command and go back to a previous state or menu, and how to correct during data entry. "Back/Cancel/correction" or similar words should work in a consistent way over different context or it must be very clear how they differ.

**System playback** If the system repeats a command back to the user as verification this should not be done until the user is finished with a complex entry, for example a long sequence of digits, such as a telephone number.

**System misrecognition** It should be clear if the ASR has failed or the user command was wrong when the system cannot carry out a command from the user. It is important to consider if and how certainty thresholds are used and if and how confirmation is used.

**Microphone on/off status** It should be clear to the user when the microphone is on. When using push to speak the system can use auditory and visual feedback to indicate that the microphone is on. It is also possible to have the user push the button during the whole utterance. Another solution is for the system to be always listening.

**Timing of microphone on** If push to speak buttons are used and there is a delay from the press of the button until the system is listening, this must be clear for the user that she needs to wait before giving a command. This also applies when the system is prompting the user for more information and wait for an answer. Another solution is to allow the user to barge in and always listen, even during system utterances.

**Pace data entry** It should be clear to the user if and how information can be entered, for example grouping of numbers in telephone numbers, or more restricted with digits followed by pauses.

**Order of data entry** Prompts and displays should be clear as to what informations is needed and in what order to complete a request. It should also be clear if there is some information that is already given, for example a state or city in



the destination entry. This can be done by more detailed prompts that indicate what information that has already been provided.

**Format of data entry** It should be clear what format is expected for information requested from the users, for example addresses with street name and number, intersections with two streets.

**Leading prompts** Prompts should clearly indicate if there are various ways to input the requested data, otherwise, if several types of data are acceptable but only one type is prompted for, users seldom uses other better or more intuitive

**Wording of commands** The vocabulary needs to be targeted to the user (and not system) using everyday words for the commands.

An evaluation of the SENECA research system confirmed some of these findings (Minker et al., 2004). In this setting the users drove a car that had the prototype system installed in the backseat. They also had a passenger that was a professional driving assessor in their car while driving. They had to perform 9 tasks that where related to destination input (5), dialling (2) and the address book (2). The tasks consisted of operations like: activating the main function, activating a sub-function, selecting an item out of a list, and input characters ('spelling') and digits.

The study showed that fewer tasks were completed in comparison to users who used a manual interface (79% compared to 90%), but the completed tasks for destination entry took significantly shorter time to complete for the speech interface (63 seconds) than the manual interface (84 seconds). The problems detected were:

**Forgotten commands** which lead to the higher number of incomplete tasks.

**Vocabulary errors** for example, wrong command for negative confirmation where users said "false" instead of "correction" and wrong commands for destination input where "indicate address" were used instead of "input address". These vocabulary errors were frequent even though the system had many synonyms.

**Dialogue flow errors** occurred when the user did not follow the predefined order, for example providing information instead of a confirmation when the system asked for the ability to provide more information

**Misrecognitions** caused problems when the user ignored misrecognitions and continued with the dialogue without correcting

**Barge in** where problematic when the user spoke while the system was still speaking.

All these errors increased the time spent on tasks and reduced the number of tasks completed. As mentioned before the completion time was still lower for the speech interface. The assessment by the professional driving assessor also showed that the number of driving errors is lower for the speech interface, significantly so for too low speed and inexact lane keeping, but there are also differences for distance to low, sudden/late breaking and no indicating. This holds especially for complex tasks where the differences where largest. The users also ranked the speech interface on a six point scale concerning safety and comfort, and it was ranked near top levels.

Another study that focused more on the experiential perspective on natural language interaction in the automotive context used a wizard of Oz approach to explore how drivers perceive speech interfaces (Wärenstål and Kronlid, 2014). The study resulted in emerging design principles ranging from high-level design values to low-level interface principles:

- Principle 1** Model driver-related tasks in three phases: pre-, during, and post-driving, and have the system assist the user in all three phases.
- Principle 2** Use sophisticated voice quality to inspire trust.
- Principle 3** Match system functionality with voice quality and sophistication, and do not raise the user expectations of the system's capabilities.
- Principle 4** Provide rich information, and motivate suggestions short and succinctly.
- Principle 5** Allow the user to shortcut the dialogue protocol, by allowing several pieces of information be inputted in one go.
- Principle 6** Ask for forgiveness, not permission. Make suggestions, and allow for correction.
- Principle 7** Use information-rich and effective language in order to reduce interaction time.
- Principle 8** Organize and name functions and domain items according to user mental models.
- Principle 9** Enable users to direct, don't force them to discuss, i.e. provide a relevant but small set of choices; don't ask open questions.
- Principle 10** Avoid unnecessary reporting, and don't report normalcy.

## References

- Tilman Becker, Nate Blaylock, Ciprian Gerstenberger, Ivana Kruijff-Korbyová, Andreas Korthauer, Manfred Pinkal, Michael Pitz, Peter Poller, and Jan Schehl. Natural and intuitive multimodal dialogue for in-car applications: The sammie system. In *ECAI 2006*, pages 612–616, 2006a.
- Tilman Becker, Peter Poller, Jan Schehl, Nate Blaylock, Ciprian Gerstenberger, and Ivana Kruijff-Korbyová. The sammie system: multimodal in-car dialogue. In *Coling ACL 2006 on Interactive presentations*, pages 57–60, 2006b.
- Niels Ole Bernsen, Andre Berton, Marcela Charfuelan, Laila Dybkjaer, Mykola Kolodnytsky, Dmytro Kupkin, and Mohamed Zakaria Kurdi. Progress report on the natural language understanding, dialogue management, response generation, and speech synthesis components. Deliverable 11, EU-project VICO, November 2002.
- Lawrence Cavedon, Fuliang Weng, Rohit Mishra, Harry Bratt, Badri Raghunathan, Hua Cheng, Hauke Schmidt, Danilo Mirkovic, Ben Bei, Heather Pon-Barry, Tobias Scheideck, Brian Lathrop, Joyce Chen, Stanley Peters, Liz Shriberg, and Carsten Bergmann. Developing a conversational in-car dialog system. In *Human Language Technology Conference of the North American Chapter of the Association of Computational Linguistics*, 2007.
- Jackie C. Chang, Annie Lien, Brian Lathrop, and Holger Hees. Usability evaluation of a volkswagen group in-vehicle speech system. In *Proceedings of the 1st International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, AutomotiveUI '09, pages 137–144, New York, NY, USA, 2009. ACM. ISBN 978-1-60558-571-0. doi: 10.1145/1620509.1620535. URL <http://doi.acm.org/10.1145/1620509.1620535>.
- P. Coletti, L. Cristoforetti, M. Matassoni, M. Omologo, P. Svaizer, P. Geutner, and F. Steffens. A speech driven in-car assistance system. In *Intelligent Vehicles Symposium, 2003. Proceedings. IEEE*, pages 622–626, June 2003. doi: 10.1109/IVS.2003.1212984.
- Nils Dahlbäck and Arne Jönsson. Dialogue systems when the dialogue is just a secondary task - some preliminaries to the development of in-car dialogue systems. In *Communication - Action - Meaning. A Festschrift to Jens Allwood*. Göteborg University, 2007.
- Dragon. Dragon drive™ connected apps and services platform – an overview. Nuance Communications, June 2013.
- Victor Ei-Wen Lo and Paul A. Green. Development and evaluation of automotive speech interfaces: Useful information from the human factors and the related literature. *International Journal of Vehicular Technology*, 2013.
- Petra Geutner, Frank Steffens, and Dietrich Manstetten. Design of the vico spoken dialogue system: Evaluation of user expectations by wizard-of-oz experiments. In *Proceedings of the Third International Conference on Language Resources and Evaluation (LREC'02)*. European Language Resources Association (ELRA), 2002. URL <http://aclweb.org/anthology/L02-1074>.
- Mukund Ghangurde. Ford sync and microsoft windows embedded automotive make digital lifestyle a reality on the road. *SAE International Journal of Passenger Cars - Electronic and Electrical Systems*, 3(2):99–105, 2010. doi: 10.4271/2010-01-2319. URL <http://saepcelec.saejournals.org/content/3/2/99.abstract>.

- John H.L. Hansen, Jay Plucienkowski and Stephen Gallant, Bryan Pellom, and Wayne Ward. “cu-move” : Robust speech processing for in- vehicle speech systems. In *ICSLP-2000: International Conference on Spoken Language Processing*, 2000.
- John H.L. Hansen, Xianxian Zhang, Murat Akbacak, Umit H. Yapanel, Bryan Pellom, Wayne Ward, and Pongtep Angkititrakul. Cu-move: Advanced in-vehicle speech systems for route navigation. In *DSP for In-Vehicle and Mobile Systems*, pages 19–45. Springer LINK, 2005.
- Paul Heisterkamp. Linguatronic product-level speech system for mercedes-benz cars. In *Proceedings of the First International Conference on Human Language Technology Research, HLT '01*, pages 1–2, Stroudsburg, PA, USA, 2001. Association for Computational Linguistics. doi: 10.3115/1072133.1072199. URL <http://dx.doi.org/10.3115/1072133.1072199>.
- Fredrik Kronlid, Jessica Villing, Alexander Berman, and Staffan Larsson. Comparing system-driven and free dialogue in in-vehicle interaction. In *Proceedings of Interspeech 2011, Florence, Italy*, 2011.
- Wolfgang Minker, Udo Haiber, Paul Heisterkamp, and Sven Scheible. Intelligent dialog overcomes speech technology limitations: The seneca example. In *Proceedings of the 8th International Conference on Intelligent User Interfaces, IUI '03*, pages 267–269, New York, NY, USA, 2003. ACM. ISBN 1-58113-586-6. doi: 10.1145/604045.604095. URL <http://doi.acm.org/10.1145/604045.604095>.
- Wolfgang Minker, Udo Haiber, Paul Heisterkamp, and Sven Scheible. The {SENECA} spoken language dialogue system. *Speech Communication*, 43(1–2):89 – 102, 2004. ISSN 0167-6393. doi: <http://dx.doi.org/10.1016/j.specom.2004.01.005>. URL <http://www.sciencedirect.com/science/article/pii/S0167639304000214>.
- Bryan Pellom, Wayne Ward, John Hansen, Ronald Cole, Kadri Hacioglu, Jianping Zhang, Xiuyang Yu, and Sameer Pradhan. University of colorado dialog systems for travel and navigation. In *Proceedings of the first international conference on Human language technology research*, pages 1–6. Association for Computational Linguistics, 2001.
- Roberto Piebaccini, Krishna Dayanidhi, Jonathan Bloom, Jean-Gut Dahan, Michael Phillips, Bryan R. Goodman, and K . Venkatesh Prasa. Multimodal conversational systems for automobiles. *Communications of the ACM*, 47(1), 2004.
- Roberto Pieraccini, Krishna Dayanidhi, Jonathan Bloom, Jean-Gui Dahan, Michael Phillips, Bryan R. Goodman, and K. Venkatesh Prasad. A multimodal conversational interface for a concept vehicle. In *INTERSPEECH*. ISCA, 2003. URL [http://www.isca-speech.org/archive/eurospeech\\\_2003/e03\\\_2233.html](http://www.isca-speech.org/archive/eurospeech\_2003/e03\_2233.html).
- Reimund Schmand. Content - driving the user experience – nuance mobile content whitepaper. Nuance Communications, August 2014.
- Toyota. *Entune™ Premium Audio with Navigation and App Suite*. [http://www.toyota.com/t3Portal/document/om-s/0M15QTQRG/pdf/15\\_EntunePremiumApps.pdf](http://www.toyota.com/t3Portal/document/om-s/0M15QTQRG/pdf/15_EntunePremiumApps.pdf), 2015.
- UConnect. *UCONNECT® VOICE COMMANDS QUICK GUIDE*. <http://www.driveuconnect.com/pdf/UconnectVoiceCommands.pdf>, 2015.

- Pontus Wärenstål and Fredrik Kronlid. Towards a user experience design framework for adaptive spoken dialogue in automotive contexts. In *Proceedings of IUI'14*,, 2014.
- Fuliang Weng, Sebastian Varges, Badri Raghunathan, Florin Ratiu, Heather Pon-Barry, Brian Lathrop, Qi Zhang, Harry Bratt, Tobias Scheideck, Kui Xu, Matthew Purver, Rohit Mishra, Annie Lien, M Raya, S Peters, Y Meng, J. Russell, L Cavendon, E Shriberg, H Schmidt, and R Prieto. Chat: A conversational helper for automotive tasks. In *INTERSPEECH 2006 - ICSLP*, 2006.
- Fuliang Weng, Baoshi Yan, Zhe Feng, Florin Ratiu, Madhuri Raya, Brian Lathrop, Annie Lien, Sebastian Varges, Rohit Mishra, Feng Lin, Matthew Purver, Harry Bratt, Yao Meng, Stanley Peters, Tobias Scheideck, Badri Raghunathan, and Zhaoxia Zhang. Chat to your destination. In *Proceedings of the 8th SIGdial Workshop on Discourse and Dialogue*, pages 79–86, 2007.
- Baoshi Yan, Fuliang Weng, Zhe Feng, Florin Ratiu, Madhuri Raya, Yao Meng, Sebastian Varges, Matthew Purver, Annie Lien, Tobias Scheideck, Badri Raghunathan, Feng Lin, Rohit Mishra, Brian Lathrop, Zhaoxia Zhang, Harry Bratt, and Stanley Peters. A conversational in-car dialog system. In *NAACL HLT Demonstration Program*, pages 23–24. Association for Computational Linguistics, 2007.

# Concept Design for Voice Control in Trucks - What do drivers actually want?

## Abstract

With trucks being one of the areas where a lot of new technology can be implemented, Volvo Trucks aims to keep their momentum on the market by integrating voice controlled features into their vehicles. This study aims to find out what truck drivers potentially could want to control by voice, and how speech technology in general could be the solution for general problems in their line of work. By completing a literature study resulting in an interview guide, two focus group discussions were held. The sessions resulted in a brief of voice features within trucks today and what drivers might want in their trucks tomorrow. The two focus group meetings laid the foundation for a field study where concepts and features were presented. Participants of the field study could then give insight in order to polish, modify or even reject previous ideas and concepts. Finally, the data gathered from the field study would then result in three major concepts and an additional feature list.

# 1. Introduction

During the last decades evolution of technology has been advancing forward in rapid pace and by doing so enabling new ways to counter previous problems in different areas of society. With trucks being one of the areas where a lot of new technology can be implemented, Volvo aims to keep their momentum on the market by integrating voice controlled features into their vehicles. Furthermore, the company wants to know what speech related features their customers and drivers actually want to have rather than throwing punches in the dark.

As a result, the purpose of this study is to build concepts of what speech technological features truck driver's could want to have in the future. This is done by first analyzing the current state of voice features in vehicles by having two focus group discussions. The first group session is held with Human Machine Interaction employees from Volvo who have experience with previously developed speech technology in trucks. The second group session consists of a group of expert drivers from Volvo's test track, Hällered. Both groups are also asked what they believe could be possible voice controlled feature for the future as their expertise makes them perfect candidates to do so. Focus group results will be the laying foundation for concept ideas that will be presented for drivers in a field study. By presenting ideas in the field study, the driver's will be able to give their thoughts on the concepts while also being able to contribute with their own ideas and needs. As there is a lot of diversity within the truck profession the field study also aims to be able to explain differences between different segments. This is a collaborative study between Volvo AB and Linköpings university, partly financed by VINNOVA, aiming to build concepts of what features truck drivers want to control by voice and how one could make their wishes come true within a near future.

## 2. Background

Human Machine Interaction (HMI) has always been an academic pillar of the evolution of cars. Initially HMI focused on how the vehicle would actually be driven. As time changed however, its focused shifted towards integrating new technological findings into the car. Damiani et al. [10] concludes the state of the art for car interiors and HMI as following:

- Increasing of the support to the driver (preventive safety).
- Satisfaction of driver's need for connectivity.
- Increasing of entertainment for for passengers.
- Re-thinking to the interior as a useful functional space.

While a truck is not the exact same thing as a car, one could argue that the two are quite resembling as they usually drive on the same roads with resembling interiors and almost the same laws and rules. The listed findings above are therefore of something that this study considers of great importance. Being able to increase the quality of driving a truck while also making it safer is something that is perhaps the highest priority of Volvo right now. Additionally, as studies have shown that speech technological features may be one of the next revolutionary steps towards safer roads, driving safely is one of the key components of this project.

As a truck driver could consider the truck's cabin as their personal work environment, the possibility to adapt the trucks interior and its technology could most probably increase the quality of their job. One way of doing this is by using a so called adaptive system. Wärnestål and Kronlid [11] defines the goal of adaptive systems as the ability " [...] to adapt interaction and interface to a specific user based on her knowledge, skills, goals, and preferences.". Further, they also add that the system is supposed to " [...] provide the right information, in the right form, at the right time in relation to the user's current state.". In addition to this, Rothock et al. [12] defines adaptive user interfaces (AUIs) as systems that adapt their available actions and visual output to the user's

specific needs at the time. The system is able to do so by monitoring the user's status but also the system state and its current situation. AUI's supposedly increases quality of use of a product by facilitating user performance, minimizing failure and therefore need to request help, easing system use, supporting users deal with complex systems and avoiding cognitive overload problems which makes it suitable for driving. From a drivers point of view, having an AUI could in theory mean the luxury of having someone working as an assistant driver by helping with what that actual driver actually needs. This leaves us with the problem of how someone driving a vehicle would communicate with a system without losing his or her focus and therefore still maintaining safe driving. Wårnestål and Kronlid emphasizes that speech has been proposed as "the ideal interaction modality" for humans [10-11,13]. Besides this, Brouwer et al. [11] suggests that speech recognition technologies could potentially aid older drivers in their performance of concurrent tasks when driving. As such, this is where speech technology and voice features comes in. It has been foreseen that adaptive speech systems is one way to increase the quality of automotive design space while still driving safe [6, 10]. In addition to this, the main advantages of using speech recognition interfaces include reduced training time, increased worker productivity, reduced secondary key input and improved timeliness and accuracy of information made available via voice [14].

Furthermore, Adriana B. and Paul G [1] had the ambition to find guidelines for usage of speech technology while driving by summarizing 15 articles on the subject. Only a few conclusions could however be drawn because of a lack of common definitions of dependent measures, unique test methodology and insufficient statistical data [9].

Besides this, it is argued that even though AUI's may have great benefits, they come at a price. The flexibility of AUI makes them hard to define and therefore makes it easy for them to violate usability design principles [9]. Additionally, speech recognition technology for instance should not be used indiscriminately. Instead it is crucial to have careful attention to the design and the complexity of the underlying system where speech recognition is going to be implemented [12] Using concepts however, one does not aim to make detailed blueprints but rather try to make a rough sketch of how a product or function would work. This study therefore has the ambition to reduce complex domains to clear models and concepts. As the strength of AUI's is to adapt themselves to a specific user's needs, building AUI concepts could be considered a complicated task. It should therefore be seen as a necessity as the diversity of such a big segment as truck drivers will result in a great lot of different needs. Additionally, voice control and speech technology in general has just recently moved into vehicles which makes it yet unknown to many drivers. By having an adaptive system that could offer different ways to solve problems, users would have the chance to adapt with the system in their on pace. The goal for said study is therefore to create concept that aims to be adaptive with the main argument being that customization within a set frame gives the driver's power to chose what to use and what not to use.

By diving into a new set of technological tools, one must be careful when designing new features as they may prove to have consequences that were not initially considered. Further, Goodwin [5] emphasizes the need for a interaction designer to think about tomorrow's design problems while analyzing today's. This means, as the the author of this article is no expert on the subject, that there is a need for iterations and greater knowledge of trucks. Going into this study, some work has already been done by Volvo however. The company had already analyzed their customer base and found seven major segments. A so called persona was created for each segment as well. Goodwin defines personas as "[...] archetypes that describe the various goals and observed behavior patterns among your potential users and customers.". As such, the project team went into the field with a good understanding of the different segments that together formed the users of Volvo trucks.

Another way of collecting expertise knowledge and insight is by using focus group interviews [9]. Krueger & Casey [8] says that the intention of focus group interviews is to make a group of people with specific attributes provide qualitative data related to a certain research topic or area in general. Moreover, a moderator is present during each session to provide guidance and make discussions easier. A focus group could therefore be seen as a specialized group with a specific



intention following a specific pattern [10, 3]. Further, it is a powerful tool to get basic knowledge and test whether ideas are even viable [9]. It is however not a good method to get detailed insight of how a user would actually use a product, but rather initial ideas. For this study this is not an issue as we want as many ideas as possible and not go on a deeper level quite yet. Two focus groups were recruited for this study and will be introduced later on.

By interviewing as a part of the field study, one will be able to understand the set of driver's goals, major tasks, mental models and as a result get opportunities for design to improve their quality of work.

## 3. Methodology

---

### 3.1 Pre Study

The pre study laid the foundation for the questions that would be discussed with the first focus group. An interview guide was initially formed in order to get answers to said questions with room for follow up if need be. The guide's questions were simply aimed to answer what drivers would want to control by voice while working. This is an important formulation as it does not limit the questions to what one could want to control inside of the truck, but rather as a part of fulfilling a job. As part of the pre study, one member was chosen to be facilitator during the focus group meetings while another would be taking notes. The facilitator in question would be giving a short introduction of both the study itself and the session in question, give follow up question if need be and also introduce new questions.

### 3.2 Focus Groups

---

In order to get knowledge about Volvos current projects that involve speech technology and the company's attitude towards it in general, a focus group was formed consisting of four (4) employees and one (1) extern consult from Semcon who all were Human Machine Interaction (HMI) experts. Being able to talk with the focus group would also result in insight about what has previously been done and give ideas for the future. The acquired information from focus group one would not only give ideas about possible concepts for further development however. It would also lay the foundation for focus group two, Volvos expert drivers from Hällered which is the company's test track for future releases and testing.

Working as a truck test driver at Hällered implies having the greatest knowledge of Volvos products as the work is based around testing the upcoming products. While the test track mainly pushes the cars driving capacity to its limit the drivers also do long haul testing cross countries. This means that even though the drivers of Hällered could be considered to work in somewhat stressful, unnatural environments that is the test track, they do not lack the experience of traveling on the road like any every day truck driver. This being said, the drivers are also working with a lot of different testing concepts which make probably makes them the best focus group one could possibly imagine for this project.

Even though there would only be one day between meeting focus group one and two, the time for iteration could prove to be of worth as the project group went in with a very open mind, not throwing away a single idea from neither session. The originally formed interview guide was still used, but ideas that came up during the first session would instead be used as followup questions.

After meeting with the two focus groups the different ideas and needs of drivers would be mapped into different categories. One reason for doing so is the simple fact that one could hope each major concept could tackle more than one problem. Further, by having different categories it would be easier to form what we would like to call concept questions. The purpose of having a concept question is that it illustrates an image of problem(s). The answer to a concept question may be simple or complex depending on the question itself. A straightforward question such as "How do I make a call without picking up my phone?" should be easier to answer in comparison to a vague question such as "Is there any way that I can get help with planning my driving route?". The vague question may however result in more than one function, and by combining more than one function one could for instance build a GUI concept answering the initial concept question. By forming a concept question, one is not only given a picture of needs but also a way to combine different features into one concept. The point of using concept questions is simply forming a question that if answered would fulfill the needs of user experience for said study. Additionally, when having a concept question one could also break down the question itself into different features, making it possible to prioritize one feature over another. By doing this it is also possible to for instance make schemes and illustrations of costs and estimated value for each feature which would make it easier to decide what to implement first or develop further.

The second reason for categorizing the gathered data is to find major themes as a result of synthesizing the categories. The purpose of the major themes would be to build scenarios from them. These scenarios would then be presented in the field study as a way to get feedback on features that were suggested in the focus group discussions but also find new one. The data from the two focus group sessions would in other words result in a second interview guide for the upcoming field study, which could be considered to be a persona hypothesis [5]. As a result of this there is a greater chance of identifying likely user roles which is important as truck drivers as a group is diverse.

---

## 3.3 Field Study

### 3.3.1 Recruitment

With the seven predefined segments that were handed over by Volvo, the initial goal was to recruit at least one participant from each segment and then add three extra drivers, making it a total of 10 sessions. Time was however limited, and it was not an easy task to work around drivers schedule to meet them in person. A decision was therefore made to do two phone interviews per lost segment. Two segments were sadly hard to get a meeting with, which resulted in a total of four phone interviews and 5 sessions in person for this study. As a thank you for the drivers participation and time, they were gifted merchandise from Linköpings university.

### 3.3.2 Pilot study & procedure

Before the actual field study started, a pilot test was held with a now retired truck driver. The pilot test focused on trying out the questions that had been formed as a result of the focus group discussions. Only small changes were made with the exception of the introduction of the study. During the pilot test the participant were taking technical limitations and problems into consideration which was not relevant for the study. As a result, each session with the truck driver in the real field study started with an explanation of the purpose of this study, and the participant was asked not to focus on what feature would be possible or almost impossible to have. Instead, they were encouraged to have an open mind and that their insight is very appreciated regardless of how vague their needs might be.

The participant was then shown pictures on an iPad using a Keynote presentation, making it possible to control it with an iPhone over bluetooth. By avoiding unnecessary distractions such as touching the iPad to swipe to the next picture, the participant would instead be able to focus on the picture shown on its screen. The purpose of the pictures was to help the participant trigger ideas

and memories by showing different scenarios relevant to the themes that had been found in the data from the focus group discussions. The technique is simply referred to as "triggers" and includes any method that helps a person to imagine something using an object or a picture. The purpose of using triggers in this study will be further explained in the upcoming analysis section. For the four phone interviews that were held, the scenarios had to be carefully described to compensate for the lack of actual trigger pictures. Other than that the procedure was not changed. After the scenarios had been presented, an open conversation was held where the driver would be able to add anything that comes to mind that had not already been covered by the interview guide.

Lastly, to compensate for the risk of taking bad notes or misunderstandings, each session was also recorded by sound after getting the participants approval of doing so. The acquired data was then ready for analysis which will be further discussed in the next segment.

### **3.3.3 Analysis**

The recorded data from each session was structured by using a model. A model could be defined as a tool that gives a description that helps people understand and communicate observed behavior. Further, the main object of modeling is to enable informed action. This is crucial as the purpose of the study is to deliver ideas that has to be well understood by the development team. A better explained concept will help the project team build a shared view of the problems, opportunities and potential next step. Goodwin [5] suggests that analysis of data should be rigorous but efficient as one should be focusing on aspects that will facilitate design and business decisions. Additionally, grounded theory (GT) argues that "all is data". This means that the data analysis should take advantage of every single bit of available information in your brain and therefore not be limited what is actually being said. By using GT with a customized model for the sole purpose of this study, we hope to be as efficient as possible without losing valuable insight from the truck drivers.

The data acquired is analyzed by using qualitative measures. As this article includes the early stages of the project, the first aim is to build a picture of driver's general user needs. By getting to know driver's and their work, it is easier to know how problems and desires can be tackled by using voice features.

### **3.3.4 Single-case analysis**

By analyzing each driver individually the focus laid upon understanding what each specific driver said. As trivial as it may sound, this is an important aspect as it is vital to understand why that specific driver thought or behaved as he or she did. Each single-case analysis was done by summarizing what had been said and explained by each driver. The results from every case will afterwards lay the foundation for a cross-case analysis driver's will be analyzed as a whole. Having a well executed interview guide is therefore important during all the cases, and as a result also having a good way to code the session is desired. Trigger pictures were therefore used for several reasons. The main purpose of trigger pictures is to help the participant imagine him- or herself in a given situation which can make it easier for thoughts and ideas to come to mind (the picture "triggers" the thought or idea). It will however also make it easier for the project team to code each case as every trigger picture represents a situation with an attached number. By taking notes for each picture or number it will be easier to find themes in the cross-case analysis as there already is a structure that is easy to follow. The pictures along with predefined questions for each scenario will therefore make up a open but controlled interview guide that will help rather than limit the session at it opens doors instead of shutting them.

As the results from two focus groups were considered to give enough knowledge to ask the right questions in the field, deductive reasoning was initially used for the case analysis. With deductive reasoning, there is an already existing hypothesis or general principle that the data is being compared to. To clarify, the two focus group discussion resulted in a bigger picture that could be broken into smaller segments with an attached hypothesis and was presented along with a trigger

picture in the field study. By using deductive reasoning previous ideas and features can either be confirmed, adjusted, further developed or even rejected during each field study session.

Inductive reasoning however aims to derive a general principle from specific data. It is however emphasized that the primary danger with inductive reasoning is getting false positives. Just because all driver's in this particular study for instance wanted to control temperature by voice does not mean that all driver's wants to do that. This is however where the charm of AUI's comes in; an adaptive system does not force features on users but rather gives them the option to decide themselves. Not a single general principle acquired needs to be rejected as it will only result in an option rather than a hard feature that the user him- or herself can decide whether to use. The false positives are still however a threat as it would be devastating to spend a lot of money on developing a voice feature that perhaps one in a thousand drivers actually want. This is should not be considered as a relevant problem at this stage of the project though as the goal is to get as many ideas as possible.

### **3.3.5 Cross-case analysis**

After the single-case analysis was completed, it was time to look at the field study as a whole. The objective of a cross-case analysis is to group and compare the individual cases to identify trends and behavioral patterns. This is mainly done by comparing individuals, but can also be done by comparing groups that can be found within the test group. In the early stages of this project Volvo provided a document consisting of seven previously defined segments of their customers and their needs. What mainly defines a segment is what said truck drivers job actually is. A driver that for instance delivering timber will be driving a lot off road in contrast to most cargo trucks that only drive on highways for instance. Further, someone that drive across the country will be sleeping in the car while others may sleep at home. The diversity of the truck profession lead to a minimum requirement to recruit at least one driver from each segment for this study. As we dove into this project with the seven already segments with an attached persona that Volvo had put together in a document much work was already done. The predefined segments would make it possible to group the personas together and find voice features that fit under one or more of these segments, making the cross-case analysis more efficient. It would also help with finding major differences between segments which was something this study hoped to do. With the newly acquired information from the field study, three major concepts were finalized along with an additional function list.

## **4. Results**

---

### **4.1 Focus groups results 1**

The results from each focus group will be summarized below, followed by a discussion comparing the two.

#### **4.1.1 Focus group 1: HMI-experts from Volvo & Semcon**

An analysis of the data from the first session showed three different themes that were discussed: (1) current state of speech technology; (2) what functions that possibly could be good to control by voice and; (3) what disadvantages speech technology could result in. These three themes will be further explained below.

#### **Current state of speech technology**

While writing this, there are neither text to speech or speech to text within trucks. There is however technology making it possible to make calls, navigate and start applications. The main focus while bringing speech technology into trucks is to decrease possible distractions and create an assistant for the driver.

### **Speech technological functions**

#### Reading

- Being able to get traffic information
- Being able to get vehicle data
- Being able to get a status check
- Being able to get coaching messages
- Being able to get warning messages

#### Voice Control

- Being able to ask when one needs to take a break and where the closest rest area is
- Being able to manage order processing and mail
- Being able to join road trains

#### Wanted effects

- Being able to control functions from different areas of the truck
- Being able to move to manual steering
- Being able to map up own functions through an adaptive system

### **Disadvantages with speech technology**

#### Current voice engine

- Listening to the voice too late
- Not being able to interrupt voice
- Not knowing what different options are being available
- Voice activation through a button

#### Error management

- When something has gone wrong, tests have shown that the user wants to stop the sound and go back to the beginning.

#### Management at complex functions

- Voice control could be bad for tasks that take long time and/or have long dialogues.

The results from the first session gave birth to ideas where speech technology was not only limited to being able to control current features by voice, but also use the power of being able to have a dialog with said system. The HMI group did however emphasize the current state of speech technology where there is no perfect voice engine. While previous concepts and prototypes have been good on paper, their results have been varied due to the fact that the voice engine simply does not work well. It is therefore vital for any upcoming project that a new engine is bought as it is the one thing you cannot strip resources from. Further, as the technological aspects of the interior of a truck generally is around five years behind a car, a lot of features that were tested before are already working as intended in cars. Even if this of course is good news, one could argue that the functions are made for cars and therefore not custom made for the needs of a truck driver.

#### **4.1.2 Focus group 2: Expert drivers from Hällered**

The same interview guide, facilitator and note taker were used for the second session with focus group 2. The insight from focus group 1 did however give a new set of follow up questions, which most probably did make the second session richer in content.

## Speech technological functions

### Reading

- Being able to get SMS and mail read aloud
- Being able to get serious/red warning messages read aloud
- Being able to get laws and rules read aloud when entering a new country

### Help

- Being able to navigate
- Being able to control functions from other areas of the truck, for instance temperature and light from the bed
- Being able to get traffic messages in foreign languages translated and also be given the possibility to be enlightened on new traffic rules when entering a new country
- Being able to have a system that could read traffic signs, translate them and then read them out loud
- Being able to get information when one needs to rest, where to stop and if there are available spots at said resting area
- Being able to have a digital manual that one could ask specific questions that could be read aloud
- Being able to have a navigation system that could warn the driver if he or she would be unable to reach his or hers destination without filling the vehicles tank and/or taking a break
- Being able to bring personal settings between different trucks, for instance seat, mirrors etc.

### Wanted effects

- Remove unnecessary buttons and such to make more space for other things
- Be able to focus more on the road
- Being able to save personal settings between different trucks
- Being able to activate voice features by pressing a button

## Disadvantages with speech technology

### Distractions

- It could be annoying with error messages read out loud
- It is frustrating with error signals making noise and repeatedly warning
- It is important that one does not have to go through a complete guide of the car when using a tutorial system

### Technology

- It could be difficult to use speech technology when there are more than one person in the truck
- Many voice engines today have great difficulty with dialects and different pronunciation in general

### Combination with other activities

- The speech technology take into consideration that truck drivers often make calls or listen to music while driving

---

## 4.2 Focus group results part 2

The first focus group session resulted in various insight both about the current status of voice integrated features but also visions about the future. While the original interview guide was used for the second session as well, knowledge and ideas from the first session was brought up during the second session as followup questions. The results from the two sessions combined lead into different needs and desires that could be tackled with voice controlled features.

After mapping out the different functions, four main themes were found:

1. Control already existing functions by voice instead of analog or digital buttons.
2. Get vocal assistance and further explanations when wanted.
3. Introduce an advanced navigation system with realtime updates that one can speak with.
4. Save personal settings into a mobile device that can connect with the truck. These settings can be called upon by using speech.

These themes laid the foundation for the field study. By converting the themes into theme questions to ask the drivers, previously defined function were confirmed to be something wanted and new features were found to answer the same questions. The questions are listed below:

1. What do you want to control by voice?
2. At what situations do you wish that you could talk with your car?
3. When planning or navigating throughout your day, is there anything you would like help with?
4. How would you like to personalize your vehicle, and what settings do you wish to have with you regardless of what truck you are driving?

The first part of each field study session aimed to get answers to the theme questions above. To make it easier for the participant to imagine different scenarios where speech technology could increase the quality of his or hers job, trigger pictures were used. The second part of the study focused on presenting more detailed functions from the main themes that had not yet been brought up by the participant him- or herself. The result of the field study was that previous found features were confirmed relevant but also polished, and that new functions were found. Theme question 1 told us what drivers did want to control by voice, but also what they did not wish to control by voice. Theme question 2-3 laid the foundation for three concepts that will be presented later. Functions that did not fit into the three concepts will be presented last.

#### 4.3.1 Differences between driving segments

As mentioned earlier, there were seven previously defined segments that Volvo had put together in a document. While one could assume that a truck driving off road in a forest needs to be more powerful than one driving on expensive roads, there seem to be two different drivers from this study's point of view - drivers whose days are very much alike and drivers whose days differ a great lot. If you for instance deliver gravel from the same place everyday to a limited and known amount of customers, you probably do not need an assistant helping you find places you know by heart. In contrast, if you do not even know where you are going to sleep tonight due to the fact that you only know your very next route, an assistant helping you plan routes could be **gold**. Further, a lot of drivers do indeed drive the same truck every day while others switch a lot which for instance makes the personalization of one or more vehicles rather complex. One key aspect of this study however was to build AUI concepts, and as the name hints the adaptivity forces the systems to be big as people want different. It is however important to take the segments into consideration when choosing which features to implement first as you want to satisfy as many customers as possible.

#### 4.3.2 General results

Before breaking down the concepts and discussing them through the point of this study's participants, it is important to clarify that the aim of this paper was to provide information for what truck drivers would want in their vehicles. While something could either look weird or good on paper, it is hard to know how concepts would work in the real world. With this said, all drivers contributing to this study had something that they really wanted which hopefully will make the following segment interesting and of worth. Besides this, there seem to be a lot of frustrations regarding the rules and laws by EU, especially the ones regulating driving times. Even though the solution according to the participants of this study would be to reform the laws completely, this is not something that Volvo as a company can do. It is however something that has been taken into

consideration when designing concepts as there still are ways to help the driver have more control over his or hers workday. On the other hand drivers seem to be quite happy with their interior and the functions that are currently available. The reason for this could be that they have what they need to lead a comfortable work life, especially the more experienced drivers. Not a single driver did however reject voice controlled features, and all of them expressed interest or desire for more than one function. So, without further ado, the concepts and functions will be presented in the next section.

---

## 5. Concepts and functions

This study resulted in three different concepts after three major iterations and lesser brainstorming sessions. While all concepts fulfill needs found in said study, not all of them are as easy to implement as others. The aim of this study was however to find ideas and desires without technical limitations, which makes the concept something to at least strive against. All presented functions have been confirmed as something that could be useful by truck drivers, regardless of when the idea initially arose.

Each concept is being presented with an introduction, a concept question and a function list where the concept has been broken down into pieces. There are however also some functions that could be interesting but did not quite make a complete concept. These will be recorded for in an extra kind of segment listed after the concepts. A simplified function list will also be included in appendix where the function will only be stated rather than argued for.

### 5.1 Concept 1

One inevitable problem with being a truck driver is that a truck will not last forever no matter how well you treat it. Some day one will find him- or herself in a new vehicle, perhaps of a new and yet unknown brand, feeling frustrated and lost. While there is a manual for the truck within the vehicle itself and tons of information on the web, there seem to be room for improvement when it comes to actually selling the product to its driver. Having satisfied customers is of course of the greatest interest regardless of the company in question, which is making this segment an interesting area to explore the usage of speech technology in.

#### 5.1.2 Concept question 1

*"I am sitting in a new truck. I cannot find the buttons that I am looking for, I am getting messages that I do not understand and I have paid a lot of money for functions that I do not even know how to use. What am I going to do?"*

#### 5.1.3 Function list, concept 1

1. Digitalize the car's manual. By storing a digital manual in the cars interface one could search through it by using words or by speech and by doing so finding information about function, error messages etc. more efficiently. The manual enables the user to use the vehicle maximally.



2. Make the car's interior smart. If the user would say "what does this button do?" and then press on one of the vehicles analog buttons, the user is going to get a short introduction of said button and get the choice to get further information if need be.
3. Help the user find in the truck as a many vehicles look different. One could avoid frustrating scenarios by letting the user say "where is the hand brake?" and thereafter either highlight the hand brake or show where it is located on the trucks screen.
4. Create a program for a fast introduction of the truck. Revenues is what make a company go around, and you get them by selling a product. A so called tutorial system could not only help a novice find in an unknown vehicle, but also promote the truck's different features, making it sell itself.
5. Enable a dialog between the user and the vehicle. If a warning symbol shows on the instrument board, give the user the possibility to ask for more information and possible solutions.
6. Let users communicate with each other. A problem based community could be an effective platform where user's can communicate with each other, giving and asking for advice. By doing this and adding the element of having Volvo experts, one could not only help customers but also get more insight of what people actually want to have in their trucks.

## 5.2 Concept 2

The without question most frustrating problem for 5 out of 7 segments is sadly something that Volvo cannot eliminate completely - EU rules. The majority of drivers expressed great frustration towards the fact that they often have to take their required break in the middle of nowhere to avoid having to pay fine for breaking the so called driving and rest act. Many drivers also expressed that their work is very stressful which makes it hard to find the time to properly prepare the day of work. As a matter of fact, some drivers would not even be able to plan their day as they do not get their next destination before they have delivered their current cargo. This resulted in concept two.

### 5.2.1 Concept question 2

*"I have just started my new job and am going to drive unfamiliar routes. During my first couple of days I have lost a great lot of time taking breaks that have not matched my driving schedule well. Is there any solution for my problem?"*

### 5.2.2 Function list, concept 2

1. Help the user plan his or her route. Warn the driver when he or she has planned a route without the factors below when it will most likely be needed:
  1. Rest areas.
  2. Places to fill gas.
  3. Unavailable roads due to weight and height.
  4. Potentially longer roadworks.
  5. Roads that tend to form queue.

By providing the driver with information for the list above, the driver can have a nicer journey as you minimize the risk for unpleasant surprises and instead strive towards an as well planned work day as possible. While the functions above are not voice controlled features themselves, they are a requirement for a good dialog between the user and the car. Say that a driver for instance simply plans a route from A to B. The truck should then warn if there is no place for a proper rest or no available petrol station and if the route is so long that he or she has to take a break by law or when the fuel tank will run empty. The user should then be able to verbally ask for options, which the truck provides. The driver should also be able to ask the vehicle for suggestions when planning a route, such as a good resting area that will not result in a major detour. The system should also be

able to prevent the driver from for instance being stuck on the road. This will be further explained below:

2. Warn the driver when needed and create a dialog with him or her. Examples of warnings and potential solutions that could be of value are listed below:
  1. Your destination is unreachable without refueling. Do you want suggestions on appropriate petrol stations?
  2. You will not reach your destination before you have to take your break. Do you want suggestions on appropriate rest areas?
  3. There has been a car accident on road X. Do you want to change your route?
  4. (The driver has just delivered heavy cargo to a destination) With the vehicles new weight there is a new route to your destination that will save X minutes. Do you want to change your route?

### 5.3 Concept 3

Being updated is key for a truck driver. The most severe problem for truck drivers, as identified in this study, is planning breaks due to time rather than actual kilometers traveled.

This being said, while the driving itself does not differ much the driver demands more comfort from the truck as the time of driving is greatly increased per day. As a result it is vital that the working environment, in this case the truck cabin, is a relaxed place where the driver feels at home. This can however be problematic as many drivers often change vehicles, making it harder to consider the truck something familiar and home like. The question one has to ask is therefore what the essence of the truck actually is, and how it could be attached to the driver rather than the vehicle itself. This is where speech technology could play quite an interesting role. By having an assistant to actually talk with, the driver would be able to bring his or hers own personal assistant that has been customized to fulfill their specific needs.

This laid the foundation for concept 3.

#### 5.3.1 Concept question 3

*"At my previous job I always drove the same truck. Today, however, I switch between three different vehicles and miss the feeling of driving something of my own. Is there any way to get back this feeling with my current situation?"*

#### 5.3.2 Function list, concept 3

1. Attach data to an external and mobile unit, like a cellphone, as it makes it possible to save information that could be of value to bring into different kind of trucks. Examples of such information is listed below:
  1. Adjustable settings within the truck such as chair- and rearview mirrors.
  2. Favorite radio stations.
  3. Saved hot spots such as rest areas, petrol stations, restaurants etc.

By doing this, the driver could say things like "adjust the seat like I want it" or "please adjust the rearview mirrors" and the assistant would do so. Further, even more personalized voice features could be used. A driver who for instance drive cross countries a lot could say "please turn on one of my favorite radio channels" and the assistant would be able to play music that the driver enjoys from where they are currently driving, making FM-radio easier to save. When also saving hot spots as mentioned above, the driver could ask the assistant if there is a saved restaurant near the current route. This could be a factor when a driver have only defined that he or she wants to drive from A to B, without saving stops to the route. By having hot spots, drivers could also share these

with each other. This would make it possible to for instance say, "have any of my friends recommended a restaurant near my current route?" and if the answer is yes, a followup question could be "please tell me more of this place".

2. Make the car feel more personal by giving the user the possibility to customize the car as necessary. Example of such functions are listed below:
  1. The car's voice.
  2. How warning messages are being presented.
  3. The appearance of the vehicles computer, menu's etc.

If you are going to have an artificial intelligent agent, you probably want the option to customize it. Do you want a female or male voice? Would you prefer an english voice over a swedish if the voice engine is better and more naturally sounding? The point is that people want different, and as such there should be as many relevant and available options as possible.

## 5.4 Additional functions

Some functions did not quite fit under the three concepts above and were not homogeneous enough to form a new one on their own. They are however of value and should not be left out of this article. As such, they will be presented in a fourth list below with some general guidelines.

### 5.4.1 Additional function list

1. Voice control should be activated in two different ways. One way is to push an analog button on the steering wheel as the driver will have his or her hands there while driving (hopefully). The second way should however be activation by speech, like Apple's "Hey Siri" followed by a command, such as "Hey Siri, what is the time?". This is a crucial feature for drivers who sleep in the car for instance and want to control features such as temperature and light without having to move around in the vehicle.
2. As an addition to the first function in this list, the driver should have the option to use voice control whenever it is needed. As vaguely as this may sound, the main guideline that should be followed is to focus on making features that one could want to control from other places than behind the wheel. This could potentially result in a great lot of features, and they will therefore be listed in the simplified function list in appendix.
3. Have the option to search for jobs. While this feature requires so much more than a simple interface within the car, one driver would like to be able to search for jobs on the go. For instance, if a driver were to deliver cargo between Stockholm and Gothenburg, he or she could send a message through the car. The function would work like this:  
"Looking for work. I am arriving at Stockholm in an hour. I have this much space left in my truck and I am going to drive to Gothenburg. Please send a message if you want me to deliver something on the way for you." The car would then transform the speech to text and send it out to a community of people interested in cargo delivery. By doing this one could maximize deliveries and therefore avoid wasting fuel on half empty trucks.
4. Introduce a more advanced status check. As different drivers care more about certain aspects of the car, one should be able to configure a customized status check where each factor will be presented verbally. The driver should also be able to chose what status factors that the truck is going to warn about when for instance running low, and also how it should be presented - either by sound, speech, text or light.
5. As driving the very heavy vehicle that a truck is, adding sensors that will give vocal feedback that the driver could interact with seems to be a safety feature that some drivers want. In addition to a reserve camera, the ability to simply ask the truck if there is something behind or in its blindspots could be of use.
6. When driving in a foreign country, help the driver with translating road signs and give the option to have that specific country's rules and laws read aloud.
7. The last feature presented in this article is something that Volvo cannot fix by themselves, but is something that would have a huge impact on their customer's driving long distances. By

using already existing technology in parking lots that monitor whether a parking space is free or not, driver's would like to be able to ask their car for resting areas where they actually can rest. This would make it possible to say "where is the closest resting area with at least 5 spots available?", and the car would respond with the location and ask the driver if he or she wants to go there.

## Limitations

This study is limited by the fact that there are not only a great lot of different people driving trucks, but as mentioned earlier that there is also variation between the driving segments that were previously defined by Volvo.

As time was limited but not crucial, the choice of doing interviews was the go to option as it could be considered the most valuable design research technique along with observation. The reason for this is that it gives a lot of qualitative data in a efficient manner if done correctly. With this said however, the amount of interview done is fairly low. One should therefore be careful when using the features that were found in this study as we do not know how many truck drivers actually want to have them. It would be rather depressing if a lot of time and money was spent on a function that got great feedback in this study, but in reality something that only 1 per 1000 drivers actually want.

Another limitation is the fact that there was a deficit of members when the field study was initiated. During the field study sessions, the facilitator also had to take the role as notetaker which is not ideal. As mentioned earlier, the sessions were also recorded but did nonetheless put more pressure on the facilitator. The impact that this limitation had on the study is uncertain, but it should however be taken into consideration for future research.

## 7. Discussion

The laying foundation of the concepts of this study were based on two focus groups session. The purpose of the focus group sessions was to get knowledge and insight from Volvo. The first session was held with Volvo HMI experts and the second consisted of the expert drivers from Hällered where Volvo is testing all their cars under development. The field study consisted of meeting 5 truck drivers and having phone interviews with another 4.

As mentioned earlier, AUI's could be considered as quite unstable as they might lose their benefits if design principles are being violated or wrongly used. While usability design principles of course should be taken into consideration even at a concept level, this study aimed to deliver as many speech technological features as possible. The only limitation for the functions that were brought into this article was therefore that at least one truck driver had thought that they wanted said function. This lead to a fair amount of possible speech technological functions, and hopefully a few of them will see the light some day. But, as mentioned earlier, one should not only make prediction's about the future but also analyze today. Furthermore, those who had previous experience with voice control in general expressed that the most annoying factor within the technology is bad response time. Having a good or at least decent voice engine therefore seems to be a requirement to even begin implementing voice interaction into trucks.

Besides this, having two focus group session that were designed to both give insight about today and also imagine possible features drivers would like to have, the study hoped to give fruitful ideas and concepts. After analyzing the data acquired from the focus group discussions we consider both sessions very successful as they complemented each other rather than bringing the exact same information and ideas to the table. The information gathered from the two focus group sessions resulted in a comprehensive interview guide that brought many ideas that could be further polished and defined as the field study went on.

The pilot study for the field session showed that there seemed to be skepticism towards voice feature in general as the technological aspects have been somewhat unstable and many times

simply not good enough as it can frustrate more than help. Even though the participants of the field study were asked not to take technical limitations into consideration during the sessions it is hard to know whether or not it had an effect, although it was not uttered aloud. Subliminal limitations may be hard to counter and take into consideration when trying to brainstorm, but we hope this was not a factor during this study even though it is impossible to know.

One feature that however is very much present already in some trucks is binding information to an extern unit as concept three in this study suggests. While adding a lot of features and binding information to a truck or a driver is very interesting, the ethical aspect to it should be considered as well. As we live in a time where personal and confidential information may be both stolen and used against persons and organizations, being clear about how the information would be stored is important in order to promote sales.

As the results however only included features and ideas that drivers had expressed that they wanted, everything presented in this study should be considered as possible seeds to plant. It is however undecided what concepts that should be prioritized for further development. While some ideas got great response, the actual cost and impact on Volvos costumer base has not been taken into consideration during this project. One possible way to get more knowledge about what drivers would want could be a survey where the simplified function list as laying foundation. Drivers could then personally rate and comment each function to see which ones are the most desirable. The consequence of this would however be that competing companies would also be able to get the function list, making research produced by this study available for anyone. As a final point we hope that the findings of this study will help bring further insight into the very complex technological segment that voice interaction is.

## 8. Conclusion

This study aimed to find out what voice control related feature truck drivers would want to have in their car. By using forming an interview guide through a literature study, two focus group discussions were held to analyze today while also making predictions about tomorrow. After analyzing the data gathered from the two sessions, major themes were found that laid the foundation for concepts and features that would be presented in a field study. By using yet another interview guide, previous defined concepts and features could be adjusted, modified or even rejected by the test participant. Further, also having trigger pictures would be found to have a positive effect as it helped with the flow of the field study sessions while also making the data analyzation easier. The study has shown that by getting in touch with people that have the right knowledge and asking the right questions, one can get a lot of insight even on a low budget and with little time. Furthermore, it has also showed that by using a qualitative approach to the data collection a fairly small sample size will still generate a lot of ideas that could be of great value. With this said however, it is important to do further testing and maybe use quantitative methods to get a better understanding of what features and/or concepts that should be implemented first.

## 9. References

- [1] Barón, A. and Green, P. Safety and Usability of Speech Interfaces for In-Vehicle Tasks while Driving: A Brief Literature Review, (2006).
- [2] Brouwer, W.H., Waterink, W., Van Wolffelaar, P.C. and Rothengatter, T. Divided attention in experienced young and older drivers: Lane tracking and visual analysis in a dynamic driving simulator. *Human Factors*, 33 (5): p. 573 - 582.
- [3] Cheng, K.W. A STUDY ON APPLYING FOCUS GROUP INTERVIEW ON EDUCATION. *Reading Improvement*. Winter2014, Vol. 51 Issue 4, p381-384. 4p, (2014).
- [4] Damiani, S., Deregibus, E., Andreoen, L. European Conference of Transport Research Institutes (ECTRI), (2009).
- [5] Goodwin, K. *Designing for the Digital Age: How to Create Human-Centered Products and Services*. Wiley Publishing, Inc. 56: p. 85-113, 201-351.
- [6] Hua, Z. and Ng, L. W. Speech Recognition Interface Design for In-Vehicle System. *Proceedings of the Second International Conference on Automotive User Interfaces and Interactive Vehicular Applications (AutomotiveUI 2010)*, November 11-12, 2010, Pittsburgh, Pennsylvania, USA, (2010).
- [7] John D. Lee, Brent Caven, Steven Haake and Brown, T.L. Speech-based Interaction with In-vehicle Computers: The Effect of Speech-based E-mail on Drivers' Attention to the Roadway. *Human Factors*, 43. 631 - 640.
- [8] Krueger, R. A., & Casey, M. A. *Focus Group: A Practical Guide for Applied Research*. Thousand Oaks, CA: Sage, (2000).
- [9] Lavie, T. and J. Meyer, Benefits and costs of adaptive user interfaces. *International Journal of Human-Computer Studies*, 2010. 68(8): p. 508-524.
- [10] McTear, M.F., Spoken dialogue technology: enabling the conversational user interface. *ACM Computing Surveys (CSUR)*, 2002. 34(1): p. 90-169.
- [11] Pontus Wärenstål and Fredrik Kronlid. Towards a user experience design framework for adaptive spoken dialogue in automotive contexts. In *Proceedings of IUI'14*, (2014).
- [12] Rothrock, L., Koubek, R., Fuchs, F., Haas, M., Salvendy, G., 2002. Review and reappraisal of adaptive interfaces: toward biologically inspired paradigms. *Theoretical Issues in Ergonomics Science* 6 (2), 157–172.
- [13] Stefanie, T., et al., Towards efficient human machine speech communication: The speech graffiti project. *ACM Trans. Speech Lang. Process.*, 2005. 2(1): p. 2.
- [14] Strass, A.R., Robillard, M., Schedler, S., Peterson, M. and Rabin, R. Speech recognition as a computer graphics input technique (Panel Session) *ACM SIGGRAPH Computer Graphics*, 16 (3).

## 10. Appendix

Simplified function list

Concept 1

1. Digitalize the car's manual.
  1. Make it possible to ask the car questions from the manual.
2. Make it possible to ask what a button does.
3. Make it possible to ask where a button is.
4. Create an introduction for the car that tells the user about its functions.
5. Make it possible to ask the car for further explanation when a warning symbol lights up.
6. Create a platform where drivers can talk and learn with each other.

Concept 2.

1. Warn the driver when he or she plans a very long route without:
  1. A planned break.
  2. A planned petrol refill.
2. Warn the driver if he or she plans a route that:
  1. Includes a road that is heavy trafficked when there is a better option available.
  2. When there is roadwork.
  3. If the vehicle is too high or heavy for a specific road within the route.
3. Warn the driver and give other options when:
  1. The destination is unreachable without refueling.
  2. The destination is unreachable without taking a break.
  3. There is a car accident on one of the route's roads.
  4. The driver has a new weight which makes it possible to use a new road or makes it impossible to use a road within the route.

Concept 3.

Attach data to an extern and mobile unit, like a cellphone, as it makes it possible to save information that could be of value to bring into different kind of trucks. Examples of such information is listed below:

1. Adjustable settings within the truck such as chair- and rearview mirrors.
2. Favorite radio stations.
3. Saved hot spots such as rest areas, petrol stations, restaurants etc.

Additional functions.

1. Activate voice control by pushing an button.
2. Activate voice control by saying a phrase (Hey Volvo..).
3. Have the option to control any button that can be pushed with voice instead. Focus on buttons that drivers want to control from both inside and outside the car (lights, volume, etc.).
4. Introduce an advanced status check. The driver should chose which factors should be read out loud by the system.
5. Give the driver the option to ask if there is something close to the car.
6. Translate signs and read new laws and rules when entering a new country or give the driver the option to call for it.
7. Make it possible to ask for available rest areas and get information about how many spots are still free.

# Speech recognition technology in trucks: potential uses and implications for visual-manual distraction

Sofia Lindvall

May 27, 2015

26 May 2015

## Contents

<b>1</b>	<b>Introduction</b>	<b>2</b>
1.1	Research question . . . . .	2
<b>2</b>	<b>Theory</b>	<b>2</b>
2.1	Cognitive load and attention . . . . .	3
2.2	Risks, accidents and distractions . . . . .	3
2.3	Speech interfaces . . . . .	4
<b>3</b>	<b>Method</b>	<b>5</b>
3.1	Goal-directed design . . . . .	5
3.2	Ethnography . . . . .	6
3.3	Data collection procedure . . . . .	6
3.4	Participants . . . . .	7
3.5	Data analysis procedure . . . . .	7
<b>4</b>	<b>Analysis</b>	<b>8</b>
4.1	Focus groups . . . . .	8
4.1.1	Human Machine Interaction department . . . . .	8
4.1.2	Test drivers in Hällered . . . . .	9
4.2	Participant observation . . . . .	11
4.3	Driver information . . . . .	11
4.4	Distractions . . . . .	14
4.5	Problems . . . . .	17
4.6	Suggested functionality . . . . .	18
<b>5</b>	<b>Discussion</b>	<b>20</b>
5.1	Result discussion . . . . .	20
5.2	Method discussion . . . . .	22
<b>6</b>	<b>Conclusions</b>	<b>22</b>



# 1 Introduction

Research has shown that many dangerous behaviours on the roads are connected with visual-manual distraction. It is therefore important to come up with ways to decrease this distraction. One way to do so is to use speech recognition whilst driving, which is getting more and more common. The positive aspect is that speech recognition can allow the drivers to keep their eyes on the road which is an important step in creating safer vehicles. Speech is also a natural way for humans to communicate and we rarely have to be trained to be able to use speech recognition. However, there are questions that need to be answered in order to implement voice control functions which increases safety and in the same time are useful for the driver. One question is how to cope with the fact that cognitive load may increase using complex and not well-designed systems. Research regarding cognitive load associated with traffic risks are not well-established and some say that cognitive load increases road safety, while others show that an increase in cognitive load increases risk in traffic. It is therefore important to consider both visual-manual and cognitive load in the design of a user-friendly speech recognition system.

Truck drivers often don't have the time to turn over to make calls, to text or plan their route, which leaves them having to do these things on the road. In order for truck drivers to perform such task required in their daily work, a speech recognition system could be a safe solution. This study is an attempt to find what the user needs of truck drivers are, and how these needs can be satisfied with a speech recognition system.

## 1.1 Research question

The following is the research question: How can speech control be utilized to meet the user needs of truck drivers and decrease driving distraction?

- What are the needs of truck drivers?
- What functions do they want to control with voice?
- What information do they want to be presented with in voice instead of text?

## 2 Theory

This section consists of a theory background. The first part is about attention and cognitive load. The second part contains information about risks, accidents and distractions in traffic, and the third part is about speech interfaces and how they can be used in vehicles.

## 2.1 Cognitive load and attention

In order to understand cognitive load and its impact on speech interfaces we need to understand how the working memory is put together. The working memory holds the activated portion of the long-time memory and moves that in and out of the short-time memory. Alan Baddeley suggest that the working memory consists of four elements: the visuospatial sketchpad, the phonological loop, the central executive and the episodic buffer. The visuospatial sketchpad holds images, the phonological loop holds inner speech for verbal comprehension and acoustic rehearsal, the central executive coordinates by deciding what information to process and how to process it. The episodic memory integrates memories from the different systems to an episodic representation. (Sternberg, 2009). One of the major assumptions of the cognitive load theory is that the working memory only has a limited amount of resources (Bannert, 2002; Ayres and Paas, 2012, Young and Stanton, 2002). When the demand exceeds the available resources, the performance of a task will degrade.

It exists several theories about attention and how this is managed when we are presented to several tasks, which is called divided attention (Sternberg, 2009). Theories have moved towards limited attentional resources, which refers to a fixed amount of attention that can be allocated according to what the situation demands (Sternberg, 2009). More recent theories claims that the theory is an oversimplification and that dividing attention is easier when the attention is divided into different modalities. However, this is a theory that also has been criticized.

Young and Stanton (2002) provided an alternative to the attention theory which they call the Malleable Attentional Resources Theory (MART). Other theories of attention claims that we have a limited amount of resources available, but MART claims that the size of available resources can change depending on the task. This could mean that reducing demand does not have to lead to an improvement in performance of a task. MART proposes that resources may shrink to accommodate the demand required by the task and that this could lead to a degradation of attention and performance in tasks that doesn't requires much demand. The consequences of reducing demand of a task could lead to a driver having difficulties handling a safety-critical event.

## 2.2 Risks, accidents and distractions

Several studies have shown that inattention plays a key role in vehicle crashes (Klauer et al., 2006; McEvoy et al., 2005; Stutts et al.; 2001, Engström, 2011). Inattention can derive from several causes of distraction, such as engagement in a secondary task, fatigue, non-specific eye glance and driving related inattention to the forward roadway (Dingus et al., 2006; Klauer et al., 2006). Stutts et al (2001) defines distraction by claiming it to occur when:

*a driver is delayed in the recognition of information needed to safely accomplish the driving task because some event, activity, object, or person within or outside the vehicle compels or induces the driver's shifting attention away from the driving task.*

According to Engström (2011) safety-critical events are more likely to happen when the driver is distracted with a task that requires visual distraction. He says that the severity of the event and the risk to be involved in an event is strongly related to time spent with the eyes of the road.

In regards to risks associated with memory loading tasks, different studies have shown different results. In some studies, memory tasks actually have shown to decrease risks, talking in a hands-free phone were found to significantly reduce the risk for a safety-critical event to occur and lane variability have been found to decrease (Engström, 2011). In the contrary, some studies have shown that cognitive load decreases the ability to detect events and objects in the periphery.

To summarize, researchers agree that distraction is one of the main reasons why accidents occurs. Distraction can depend on different causes, but the distraction causing most accidents is visual distraction. The role of cognitive load in accidents is not conclusive, it seems like cognitive load may increase the ability to perform some driving tasks and decrease the ability to perform others.

### **2.3 Speech interfaces**

Speech interfaces have some clear advantages in comparison with visual-manual interfaces, especially in automotive domains. According to Ei-Wen Lo and Green (2013) the level of distraction is lower when using a speech interface and the speed for task completion is quicker in some tasks such as entering an address while navigating. They also claims that speech interfaces makes the driver keep lanes better, shortens the reaction time of events happening in the periphery and results in the driver glancing away from the road fewer times.

However, Reimer and Mehler (2013) conducted an on the road study where they found that voice interfaces not necessary are free from visual-manual demands. Their findings show that implementations of voice interfaces can be multi-modal resulting in visual-manual demands. For example, the interface might require the user to view the screen several times to be able to add information.

A positive aspect with speech interfaces is that they are promising in regards with the workload they demand, and some studies have shown that subjective workload was less with a speech interface compared with a visual-manual interface (Ei-Wen Lo & Green, 2013). Even though the subjective workload was less in some studies, an issue with other speech interfaces could perhaps be the level of cognitive load the interface is demanding and the issue of some tasks taking longer time based on speech recognition accuracy (Ei-Wen Lo & Green, 2013).

#### **Design guidelines for speech interfaces**

These findings show the importance of design principles for speech interfaces, especially in a domain where safety is the main priority.

An alternative for a user friendly speech recognition system is a system that adapts according to the users goals, knowledge, skills and preferences in a specific state of the user (Wärnestål & Kronlid, 2014, Reeves et al., 2004, Oviatt et al., 2009). Wärnestål and Kronlid (2014) conducted a study where they used a wizard-of-oz prototype to let drivers interact with a speech recognition system. They recorded the dialogue and made an analysis of the data with the result of ten design principle for speech recognition systems. They divided the design principles in four categories which they called elegance, levels of expertise, posture and harmonious interactions. Elegance refers to the quality of voice that should be sophisticated and, according to them, should match the quality of the system. They highlighted that speech interfaces needs a design that enables users to understand limitations and capabilities (Wärnestål & Kronlid, 2014).

Hua and Ng (2010) found four design guidelines when they examined speech recognition interface design for in-vehicle systems. They showed it to be important to think about the hierarchy structure which they thought should be broad and shallow, depending on the difficulties in navigating a speech recognition interface compared with a visual interface. Available options should be at top hierarchy level and menu paths should be short with an ideal numbers of levels in a menu not more than three. They should also get visual feedback and memory aids. Regarding tasks that often are executed, it is preferred to have them available with vocal short cuts to make the system more efficient. Sometimes it can be a good idea to let some functions be steered with hard keys, in cases where the output might intervene with the input or when the user need immediate control.

Reeves et al. (2004) added the importance of giving the user its own choice to choose a non-speech alternative to preserve the integrity of the user.

## 3 Method

This section gives an overview of the method of choice and the procedure for the study.

### 3.1 Goal-directed design

Goal-Directed Design is an approach to design which focuses on the achievement of goals (Goodwin, 2007). In order to create a successful design, when adding a feature one should ask whether it is helping the user accomplish his or her goals. According to Goodwin (2007) individual interviews conducted in the user environment is the best way to get information about the user needs. This will allow the user access to memory cues from the environment while answering questions, while at the same time allowing the interviewer to see cues in the environment showing what issues the user may have. Interviews are a good way of seeing the participant's view of the world and in order to do so, it is appropriate to use ethnographic techniques (Goodwin, 2007). When the data is collected the analysis of the data can begin. According to Goodwin (2007) the first step of analysing data

is to start with a single-case analysis, meaning that one should focus on one individual at a time to understand why that individual behaved as she or he did. After that a cross-case analysis is preferred, where individuals are compared with each other to find patterns. The patterns that are identified can be used to create personas. Personas will assist in several stages in the design process by showing who the users are and what the users needs (Goodwin, 2007).

### **3.2 Ethnography**

An ethnographer is interested in examining shared patterns in a culture-sharing group of people (Harris, 1968). A culture includes what the members of the group do, what they say, a potential gap between what they do and ought to do, and the artefacts they use. According to Hammersley and Atkinson (2007) participant observation and informal conversations usually are the most common ways of collecting data. However, collecting data can be made in several ways, for example through interviews, observations, documents etc. Data is normally collected in the group's natural environment, in the field, rather than in constructed situations. After data has been collected, the researcher proceeds to analyse it by providing an overall demographical description of the group and a theme analysis. This means that the researcher organizes data into meaningful themes and categories. Data can go under several categories and there are no strict rules on how to do this. When looking for patterns, this will depend on the research focus. An important aspect about ethnography is that the researcher should let the collected data itself show how it should be categorized, and not interpret the data based on categories defined before the data collection. (Hammersley & Atkinson, 2007).

### **3.3 Data collection procedure**

Two focus groups were conducted as a first step to get information about Volvo's perspective on voice control in trucks - the first focus group was held with personnel working with Human-Machine Interaction (HMI) and the second group with test drivers. After that, participant observation was conducted with truck drivers on one of their normal day's work out driving.

#### **Focus groups**

A question guide was prepared with questions covering two themes that were considered important to discuss during the interview. These themes were mainly about Volvos goals with a voice control system and about the future of trucks. The focus groups were held by two people, one led the discussion as a facilitator and the other took notes. The focus group began with an introduction of the subject and the purpose. An opening question was introduced and the group were asked to answer and discuss with each other. When needed, the facilitator asked follow-up questions or introduced a new subject.

## **Participant observation**

Participant observation was conducted in the truck cab whilst on the road driving. Conducting a participant observation enabled triangulation to validate collected data, as information was collected both with interview questions and observations. An observation guide with guidelines was prepared. The truck drivers were asked to proceed with work as usual and they were given a question if anything was unclear about the process. A semi-structured interview was conducted during the driving session. The participant was asked about his or her goals and needs in their work, questions regarding attitudes towards voice interaction and earlier experience with voice interactions. Notes were taken during the entire session and sessions were recorded when possible.

## **3.4 Participants**

Different user roles represent different needs and goals. Therefore, participant recruiting should be based on user roles (Goodwin, 2007). In accordance with this principle, the recruited participants for the participant observation and interview sessions were truck drivers that represented different work segments defined by Volvo Group Trucks. The work segments were: city distribution, regional distribution, interregional distribution, demanding long haul, light construction, heavy construction and heavy transport. Eight truck drivers participated in the study. Five occasions consisted of participant observation and three were phone interviews. Difficulties with finding a long haul driver and a heavy transport driver led to phone interviews with two long haul drivers and one heavy transport driver. A participant observation was conducted with an additional driver from the city distribution segment.

## **3.5 Data analysis procedure**

The first step in the data analysis was to read through the notes to get an overview of the material, this was the case for the data collected both from the focus groups and the participant observation. Every section in the data was assigned a code and categories were identified from patterns in the coding. The categories were based on what the data presented but also on the research question. When the first categories were identified a new round was conducted to further categorize the data in even more specific categories. Regarding the data from the participant observation in specific, data from each participant was read through and the coding began by finding categories for every individual, a single-case analysis, where a category was assigned to every section in the material. The next step was to identify appropriate voice interaction functions to solve the user needs of the truck drivers, which had to be done taking into account relevant theory.

## 4 Analysis

There are two perspectives that needs to be considered when designing a speech recognition system: Volvo's perspective and goal of implementing the system in trucks and the truck drivers wishes and needs. The first part presents the analysis of the focus groups which will give an understanding of Volvo's view. The second part contains the analysis of the participant observations and the phone interviews which gives an understanding of the truck drivers.

### 4.1 Focus groups

The two focus groups conducted were analysed using a thematic analysis explained in the method section.

#### 4.1.1 Human Machine Interaction department

The analysis of this focus group showed three themes that were discussed: (1) the current use of voice control, (2) which functions that are believed to be preferable to control by voice and (3) disadvantages using voice control in trucks.

##### **Current use of voice control**

There exists either text-to-speech or speech-to-text today. It is possible to call, navigate and start applications. The focus for introducing voice control in trucks is to decrease distractions and create an assistance for the truck driver.

##### **Voice control functions**

The voice control functions that were suggested, and the effects that are wished to be accomplished with them is presented in Table 1 below.

Table 1: Voice control functions

No.	Information to be read out loud
1.	Traffic information
2.	Vehicle information
3.	Status checks
4.	Coaching messages
5.	Warnings messages
Voice commands	
6.	Ask about break times and spots to stay at
7.	Handle orders and e-mails
8.	Connect to road trains
Desired effects	
9.	Control functions from other parts of the truck
10.	Being able to switch back to manual steering
11.	An adaptive system where the driver can map up own functions

### Disadvantages with voice control

Table 2 below presents disadvantages discussed during the focus group.

Table 2: Disadvantages with voice control functions

No.	Current voice motor
1.	Listens to the voice to late
2.	Cannot disrupt it
3.	User is not aware of available options
4.	Voice control activates by pressing a button
Error handling	
7.	When an error occurs, users wants to stop the sound and start over
Handling complex functions	
8.	Voice control might not be suitable for tasks requiring a long dialog

#### 4.1.2 Test drivers in Hällered

The same analysis were used on the collected data for this focus group. The analysis of this focus group showed two themes that were discussed: voice control functions and disadvantages with voice control.



## Voice control functions

Table 3 below presents voice control functions that were suggested during the focus group.

Table 3: Voice control functions

No.	Information to be read out loud
1.	Text messages and e-mails
2.	Red warnings
3.	Rules and laws when crossing land borders
Assistance	
4.	Control functions from other parts of the trucks, for example temperature and lightening from the bed
5.	Traffic messages in other languages translates
6.	Traffic signs translated and read out loud
7.	Get information about resting times – when they need to stop, where and if it has available parking
8.	Ask the trucks digital handbook questions and get answers read out loud
9.	Navigation system helping with planning routes based on the trucks weight
10.	Navigation system warning the driver if not reaching the destination without refuel or taking a break
Desired effects	
12.	Remove buttons for more space
13.	Focus on the road
14.	Save personal settings between trucks
15.	Activate voice control by pressing a button

## Disadvantages with voice control

Table 4 below presents disadvantages discussed during the focus group.

Table 4: Disadvantages with voice control functions

No.	Distractions
1.	Error messages and warnings read out load is frustrating
Technology	
2.	Difficult with voice control when it's more than one in the truck
3.	The voice control they use today do not understand dialects
Combination with other activities	
4.	Must be able to combine with speaking on the phone while driving
5.	Must be able to combine with having music in the truck cab

## 4.2 Participant observation

The following main categories were identified during the analysis:

**Driver information** Three personas created in purpose of providing an understanding of the truck drivers, including working situation and attitudes.

**Distractions** Visual-manual distractions that the truck drivers were subjected to.

**Problems** Problems encountered by the drivers during their daily work.

**Suggested functionality** Voice control functionality explicitly suggested by the truck drivers.

## 4.3 Driver information

This section contains three personas created from behaviour patterns in the collected data: Niklas, Lars and Jimmy. They were created to give an understanding of the truck drivers, including for example working situation and attitudes. These can be referred to in future work.

### Persona 1: Niklas



Figure 1: Niklas

#### Goals

- Execute his job in a correct manner
- Work efficiently to finish deliveries in time

*“I like when it’s stressful because it makes the days go faster”*

Niklas is 28 years old and recently started working as a truck driver. He lives with his partner in an apartment in Linköping. He works with distributing groceries and this is his first full-time job. He has worked there for three months, but all together as a truck driver for a year. He really likes his job. The routes he drives are the same every week, one is in the city and the other route is to smaller municipalities around Linköping. He knows the routes by now, and doesn’t need to use a navigator unless he has a new customer. He uses Google Maps on his smartphone if he needs assistance with navigating. However, he sometimes needs to call customers and colleges if he can’t

find the way. Even though he has the address in Google maps, it is not always clear on where to park the truck or where the goods should be unloaded. He does not like to make a mistake while working because it will take time for him to correct it. Sometimes he has forgotten to unload a package or unloaded the package on the wrong place. He is much more careful now, making sure that everything is correct. Work can sometimes be stressful but he thinks it is good because it makes the day go fast.

“I believe voice control in the truck would make my job easier”

Niklas likes technology and he thinks it is fun to try new ones. Sometimes he gets frustrated when he doesn't understand, but usually he learns fast. He would enjoy having voice control in the truck and thinks it would ease his job, and also make it safer as he would not have to look away from the road as much. He usually never reads text messages while driving, but he sometimes glances at the phone to see if it is anything important. Getting warnings and information in voice is nothing that he thinks would bother him at all. However, he believes it would be annoying if there was no way of turning it off. He uses social media such as Facebook, but only on his breaks. Normally there are no problems when he is out working. He has a good routine on where to stop and take breaks and where to drive. Niklas spend a lot time on the road listening to music and talking on the phone.

## Persona 2: Lars



Figure 2: Lars

### Goals

- Find a way to make his days less stressful

*“There is nothing worse than warnings and sound in the truck”*

Lars is 38 years old. He lives in a small municipality and needs to commute to work. He

lives together with his partner and they have grown-up kids. He drives long-distance traffic within Sweden and is out driving all week and, thus, gets home in the weekends. During his career he has gotten the experience of driving other types of jobs, such as city driving, gravel and package delivery. He has worked as a truck driver for 18 years.

“Stress, stress and stress”

Lars never knows what his route will be before the day comes. The destination is often in new places and he needs to use a navigator. Sometimes he uses it like a planning tool, as it shows how long the journey is expected to take. He thinks that his job is extremely stressful and he almost never has the time to sit down and take a break. He thinks it is getting more and more stressful, compared with when he started. When he does have time to take a break, there is almost never a place to stay at within reach. He does not have time to drive around and search for a place; he needs to find stops that fits with his break schedule and is near the route he’s on. Technology is something he enjoys if it works and is easy to use. His navigator works just fine, however he turns the volume down so that he does not get disturbed by the sound. He thinks it is enough just seeing the driving instructions on the screen. Another reason is that he often speaks on the phone and he believes another sound would bother him. He does not have any experience with voice control, but can imagine it would be helpful in some situations, like when it is a lot of distraction on the road. It is important that it does not intervene with him talking on the phone or when he’s got a passenger in the truck. There also needs to be a possibility to choose whether to use it or not, but if it requires a lot of settings he would never have the time or energy to use it. Warning sound that have important implications could be okay, but then only in very serious situations. In general, he thinks there are way too many warning signs and he ignores most of them.

### Persona 3: Jimmy



Figure 3: Jimmy

#### Goals

- To avoid problems that could affect his earnings

*“I dont like technology, you are supposed to think for yourself”*

Jimmy is 44 years old and has been a truck driver for 7 years. He lives by himself in an apartment and is working a lot. He has his own company and drives everything from distribution, timber, gravel and waste. He enjoys working as a truck driver for most of the time and he likes the variation, even though it can be quite stressful sometimes. He thinks having his own company is stressful because he always need to think about earning money. Some assignments are paid by commission, and if he gets upheld by for example a road block he knows that he will earn less money that day. Because of the variation in routes, he needs to use a navigator. He uses Google maps on his smartphone. He has no time for stopping to make a phone call, send a text message or adding an address to the navigator. Sometimes he feels that the risk increases when he uses his smartphone while driving, but he thinks he is quite safe anyway. He sometimes listens to the radio. When he speaks in the phone it is mostly for making work related calls. He feels a frustration about technology which is too complex. He has no patience with technology that does not work and feels that it is always on the verge of breaking. He does not trust the navigator entirely so sometimes he uses a map book. He has no need or desire to use voice interaction in the truck, because he believes it would never work smoothly. He has never tried it before. Getting any information or warning in sound would probably be very annoying and it works well as it is today.

#### 4.4 Distractions

Several visual-manual distractions where identified and these with be presented in this section.

## **Mobile phones**

The most essential function, and maybe the most obvious one, needed to avoid visual-manual distraction is to enable truck drivers to easily connect their private and work phones with the truck's infotainment system, allowing for the truck drivers to interact with their phone via voice commands. A great part of many truck drivers days consists of communicating with colleges, customers, family and friends. Speaking with friends and family is not essential, but the communication with colleges and customers is needed for them to be able to achieve their work tasks. This includes making calls, answering calls, using music applications, sending text messages and emails, and getting text messages and emails read out loud. All of the truck drivers in the study frequently used their phones as a part of their work tasks. All observed truck drivers were subjected to visual-manual distraction when using their mobile phones while driving. This was needed for them to perform their job.

## **Navigation**

Six of the participants in the study needed navigation aids in some situations (two of them never did): three of them used the Google maps application on their mobile phones; two used a separate GPS. No one had an integrated GPS. One of them double-checked with a map book, as he did not trust the GPS. With the perspective of avoiding visual-manual distraction, adding addresses and searching for addresses via voice is essential. Visual-manual distraction generated by navigation issues were seen several times during the observations. The fact that the route can change during their day makes it more difficult to plan ahead and if they dont have time to stop to modify the GPS, voice commands would prevent the visual-manual distraction.

## **Writing information down**

Another issue that came up was that the drivers had a need to write information down. With a flexible job, as was the case for most of them, they could get a phone call during the day with a new assignment. Information from the call often needed to be noted as a memory cue. Two truck drivers needed to write information down when driving, and they solved this by having paper and pen located next to the driver's seat. Searching for a pen and a piece of paper during or after a phone call to write down information led to the drivers taking their eyes of the road. Enabling drivers to note information with voice could be a way of reducing visual-manual load in this case.

## **Handling of buttons**

The observation showed that most buttons where easy to use while driving because they were located close to the driver on the right hand side of the driving seat. However, some buttons were expressed to be difficult to understand, and before the driver is familiar with a truck there can

be issues in regards to where they are specifically located. Knowing exactly where the button is located enables the driver to interact with it without taking his or her eyes off the road. Before this knowledge is established, the driver might have to glance at it while driving and also spend time figuring out what it does. Reading in the handbook while driving to get information about where a button is located and/or what it does, causes create visual distraction. A means of avoiding this could be to let the driver ask the system about the function of a button. However, explaining the appearance of an icon is difficult. For buttons easily pressed from the driver's seat, an action such as holding the button in for an amount of time to get voice information about what the button does, could be an alternative.

### **Functions not reachable**

Functions not reachable from the drivers seat that still need to be assessed while driving could be, if not physically reached, reachable with voice instead. An example that came up during the study was that one truck driver had issues with the sunblock in the front windshield, which he could not modify while driving. If the driver is not able to get the sunblock down when there's strong sunshine, this could be a safety issue.

### **Artefacts in the truck cab**

The drivers working with deliveries had packing slips in the truck cab. These packaging slips had an additional function to the obvious one (that they are to be handed over to the customer on delivery), which was that they often constituted a memory cue. The slips were sorted in the delivery order, which means that the driver could easily glimpse the pile of packing slips to see what the next stop is and what the address to that stop is. However, the packaging slips were frequently referred to and did indeed create visual distraction for the drivers. There were cases where the drivers had their phone in one hand, the other hand on the wheel and their eyes on the packing slip. These situations were necessary because the driver needed to add the customer's address written on the packing slip to the navigation application in their phone, and they had no time or no appropriate place to stop at. Keeping the addresses stored and enabling the drivers to access them by interacting with the navigation system with voice would prevent these situations from happening. If the driver could ask for information about what the next stop is, the pile of packaging slips might not be needed as a memory cue. Consequently, a potential visual distraction would be eliminated.

### **Summary**

The following table 5 shows an overview of the distractions that have been identified, how many drivers that were exposed to the distraction and how this could be solved with a voice control function instead.

Table 5: Distractions and voice control solutions

<b>Distraction</b>	<b>Frequency</b>	<b>Voice control solution</b>
Text messages and phone calls	8	Connect phone to infotainment unit and enable voice interaction
Navigation (smartphone applications and GPS)	4	Connect phone to infotainment unit Navigator with speech recognition
Writing information down	3	Adding notes with voice interaction
Handling of buttons	2	Information about functions of buttons in voice Non-reachable buttons controlled with voice
Functions not reachable	2	Control with voice
Artefacts in the truck cab	6	Integrate as much information as possible so that the driver can use voice interaction instead

## 4.5 Problems

This section is an overview of different problems that the drivers in this study encountered during their daily work.

### Remembering activities

The driver card that saves information about the driver history was told to create problems in some situations, such as remembering to add the starting country or noting when they take a break. The rules for driver times and breaks are something that of course cannot be manipulated. However, a different solution with voice could be used to assist the driver in some problematic driver situations caused with the presence of a driver card. As the driver needs to insert the card before driving away, forgetting to do so could easily be done without any kind of memory cue. During observation one of the drivers drove away before remembering to insert the card. If he would have received some kind of reminder, for example a voice reminding him about the card, that situation would not have happened. When starting the truck there is a chance that several warning signals are showing on the display (because of the status check), which is an argument that the warning for inserting the driver card should be done with sound over for example a visual warning. The warning only needs to be activated once a day, which makes it less prone to cause frustration.

There are more actions related to the driver card during the day. The driver needs to put in the country in which they are starting their journey. They also need to signal every time they take a break, otherwise the driver card will show that the driver did not take a break. In the cases covered by this study, these things were not warned for either.



Another activity, also involving forgetting, that came up during the observations was when a driver shut down the engine and tried to start it again, without success. It took a while before the driver figured out that it was the alco-lock that needed to be addressed. Assistance with figuring this out could save the driver some time. The assistance could be made with a voice telling the driver what the issue is.

## Navigation

Another issue the driver card implicitly creates is that the drivers have to plan their journey so that they can take breaks in appropriate times and in places where they actually can take breaks. Several drivers expressed the problem of finding spots to stay at where they could both purchase food and park the truck, adding the fact that they need to find such spots within a specific time slot does not make this task easier. This problem could be solved with a navigation system taking into account driver times and which finds break stops when the drivers wants and needs to take a break.

Saving time and in that way earning money is central for some truck drivers, especially those that have their own company or drives for commission. If something upholds them, such as a road block, it can make a negative impact. A navigation system that can keep track of road status and make a new journey plan based on this could be of great help in some cases, and for this end voice interaction technology is needed (see section 3.2).

## Summary

The following Table 6 shows an overview of the problems that have been identified, how many drivers that were exposed to the problem and how this could be solved with a voice control function instead.

Table 6: Problems and voice control solutions

<b>Problems</b>	<b>Frequency</b>	<b>Voice control solution</b>
Remembering activities	7	Voice reminders
Navigation	3	Enable driver to interact with navigation system with voice to get information about break times and spots to stay at

## 4.6 Suggested functionality

Except from the suggested functionality above, which is based on needs observed during the study, the truck drivers explicitly expressed several suggestions of their own, which this section will present.

This section will not in detail explain functions that have already been addressed in the section above, even though they did report them: make phone calls, answer phone calls, send and receive text messages, etc.

### **Navigation**

Navigation assistance can be helpful both for those who drives the same route every week, and for those who drive to new places. Truck drivers with experience of driving abroad expressed that a system identifying the meaning of foreign road signs and reading the meaning of them out loud would make working and navigating abroad easier.

One of the drivers driving the same routes every week said that it resulted in him knowing exactly where to drive. He also said, however, that this sometimes made him so relaxed, causing him to miss making a turn to get to his destination. Adding locations, such as break stops or customers, and reminding the truck driver when getting close could be a good solution to this.

### **Handbook**

The handbook contains a lot of information and is difficult to refer to while driving. Integrating the handbook with the truck's computer and enabling interaction with voice could save time and decrease visual-manual distraction. Two of the drivers expressed that this would be of great help when new to a truck.

### **Warnings**

One of the truck drivers said that he wanted all warnings in speech. Another said that warnings that are difficult to understand can be explained further by speech. Four expressed that they did not want warnings to be expressed with speech at all. However, they did all quite agree that red warnings, warnings that are serious and need to be followed by immediate action, could be expressed in speech.

### **Status checks**

Be able to ask about road status when driving, including information such as how long you have driven or what the status of the road is further down, will make it easier to plan the journey ahead.

### **Summary**

The following Table 7 shows an overview of the areas where driver's wished for voice control, how many drivers that mentioned it and the actual suggestions of voice control function given by the drivers.

Table 7: The truck drivers voice control solutions

Truck driver suggestion	Frequency	Voice control solutions
Navigation	8	System that translates foreign road signs and reads the meaning out load Voice reminders when close to added locations
Handbook	3	Voice interaction with the trucks handbook
Warnings	4	Get warning difficult to understand explained in voice Get red warning read out load
Status checks	4	Be able to ask about road status

## 5 Discussion

The section divides into a discussion of the study results and a discussion about the method.

### 5.1 Result discussion

The main goal of implementing voice control in trucks is to avoid visual-manual distraction. With the background that a safety-critical event is more likely to occur when the driver is subjected to visual distraction (Engström, 2011) and the results from this study showing that truck drivers repeatedly are performing task that creates visual distraction, a speech recognition system in trucks seems like a good idea as the level of distraction is lower with a speech interface (Ei-Wen Lo and Green, 2013). It will allow the user giving commands and receiving information without viewing the infotainment unit. The results have also shown that the tasks that creates visual-manual distraction is essential in the truck drivers daily work. This points to the importance of creating ways of letting the truck drivers perform these task without the visual-manual distraction it is currently causing. Implementing easy to use voice control in the truck could be a useful solution. With the main goal of decreasing visual-manual distraction, the section that describes identified distractions is the most important to consider. The frequency of how many of the truck drivers that were subjected to the distraction can also be used as a guidance regarding the importance of considering a voice control solution satisfying the need.

The other goal of implementing voice control in trucks, as discussed in the first focus group, is to create an assistance to the truck driver. The results from this study have shown several situations where the truck drivers are in need of assistance, and with the focus of decreasing visual-manual distraction voice control is the best solution. The problem section is not to be seen as essential in

regards to safety. However, implementing voice control solutions helping with the identified problems would be helpful for the drivers and a way to utilize the needs of the truck drivers. It is not always possible to meet the request from users. However, the suggestions explicitly expressed from the truck drivers can guide to an understanding of what issues the truck drivers think of, and the way they think about voice control in trucks.

A speech recognition system would enable the driver both receiving information and giving commands. However, several of the truck drivers in this study expressed the importance of being able to choose whether to use speech or a non-speech alternative. According to the design principles that were described in the theory section, it is important to give the driver the choice to use speech or non-speech (Reeves et al., 2004). From the perspective of the truck drivers in this study, it is preferred having to press a button or in some way activate the recognition before giving a command. The reason for this is because the truck drivers often perform activities in the truck cab that not always are compatible with a speech recognition system, such as talking on the phone or listening to music. When implementing a speech interface in the truck, this is important to consider.

The idea of giving truck drivers spoken commands in form of warnings or other information in sound has both its advantages and disadvantages. The strongest disadvantage that was identified in this study is the truck drivers attitudes towards getting those kind of warnings. However, 4 out of 8 participants wanted red warnings or warnings difficult to understand spoken out loud. Red warnings is the same for all truck drivers and rarely occurs, which is an argument for giving these warnings in sound. However, regarding warnings difficult to understand this might best be solved by giving the user the option to ask the system to further explain what a warning that occurs on the display means, if needed.

Navigation was something that several of the participants in this study used. Navigation with a speech interface while driving has been shown to be quicker than when manually adding the address (Ei-Wen Lo and Green, 2013) and it is decreasing the visual distraction as long as the speech interface is well-designed and does not required the user to for example view the screen several times (Reimer & Mehlers, 2013). The results from this study showed that navigation can be used in several ways, for example by showing the way to a new location, giving reminders when the truck is close to a destination or as a planning tool regarding breaks and driver times.

An aspect to consider when implementing voice control in truck is the goals, attitudes and opinion the target group is expressing. This study has provided an insight to this through the personas. The personas created can be referred to in the design process when evaluating if the function would be useful for the target group and give an initial understanding of how they might react to it. For example, when considering an adaptive interface it is easy to see that it would fit Jimmy, as he has no time for changing settings and is only interested in having the most important functions available.

The results from this study gives an understanding of what speech recognition functions that could be helpful for truck drivers. It also has the ambition to present Volvo's current view of the role of a speech recognition system in trucks. Many of the functions that were presented during the focus group came up during the participant observation as well. However, the participant observation added knowledge about the truck drivers and their situation giving an understanding of what speech recognition functions that are most important for them.

## 5.2 Method discussion

The method was constructed to enable triangulation, which in this case means that participant observation whilst the truck drivers were on the road driving would validate the collected information by both interviews and observation. However, during the recruitment difficulties with finding long haul drivers came up. The long haul drivers that were interested in participating could not plan their days ahead, and only knew where they were going the same day. Their route could also be anywhere from north to south Sweden. In order to solve this and receive information from them anyway, phone interviews were conducted. The phone interviews led to collected information about the driver and their attitudes toward speech recognition, which were valuable. However, no information regarding distractions they were subjected to were collected. The participant observation gave examples on incidents that caused visual-manual distraction, sometimes incidents that the driver had not been thinking of before. The observations could also validate that the information given by the drivers where grounded in how they behaved, as people sometimes are not aware about activities they perform. The preferred would have been to have participant observation will all truck drivers in the study. Therefore additional occasions with observation should be added in the future.

The interviews were conducted so that some of the questions were about the participants opinion about speech recognition. However, it can be difficult to know how to react on a system before being presented to it. It is therefore important that future prototypes are tested on truck drivers because the attitudes and suggestions the truck drivers expressed during this study could change when actually trying voice control in the truck. Functions that are thought to be implemented to increase safety should not be dismissed referring to the results from this study, as safety is the main goal. The results from this study shows the user needs of eight truck drivers. In order to be able to rank the urgency of implementing functions, user needs from additional truck drivers need to be taken into consideration.

## 6 Conclusions

The goals of implementing a speech interface in trucks are to decrease driving distraction and meet the user needs of truck drivers. This report contains several suggestions of voice interaction

functions based on a study made with eight truck drivers and two focus groups with test drivers and experts on Human Machine Interaction in Volvo trucks. The results from this study should be seen as an incomplete step toward fully understanding the user needs of truck drivers. The possibility of the technical functions suggested has not been taken into consideration. More research on the subject of cognitive load and safety on the road should be conducted to be able to further speak on how a speech interface should be designed and what functions it should include.

## References

- Ayres, P., & Paas, F. (2012). Cognitive Load Theory: New Directions and Challenges. *Applied Cognitive Psychology*, 26(6), 827-832. doi:10.1002/acp.2882
- Bannert, M. (2002). Managing Cognitive Load—Recent Trends in Cognitive Load Theory. Commentary. *Learning And Instruction*, 12(1), 139-46.
- Engström, J. (2011). Understanding attention selection in driving: From limited capacity to adaptive behaviour. Chalmers University of Technology.
- Goodwin, K. (2009). *Designing for the digital age : how to create human-centered products and services*. Indianapolis, IN : Wiley Pub., cop.
- Hammersley, M., & Atkinson, P. (2007). *Ethnography. [Elektronisk resurs]: principles in practice*. London : Routledge.
- Harris, M. (1968). *The rise of anthropological theory: A history of theories of culture*. New York: T. Y. Crowell.
- Hua, Z., & Ng, W. L. (2010, November). Speech recognition interface design for in-vehicle system. In *Proceedings of the 2nd International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (pp. 29-33). ACM.
- Klauer, S. G., Dingus, T. A., Neale, V. L., Sudweeks, J. D., & Ramsey, D. J. (2006). The impact of driver inattention on near-crash/crash risk: An analysis using the 100-car naturalistic driving study data (No. HS-810 594).
- Lo, V. E. W., & Green, P. A. (2013). Development and evaluation of automotive speech interfaces: useful information from the human factors and the related literature. *International Journal of Vehicular Technology*, 2013.
- McEvoy, S. P., Stevenson, M. R., McCartt, A. T., Woodward, M., Haworth, C., Palamara, P., & Cercarelli, R. (2005). Role of mobile phones in motor vehicle crashes resulting in hospital attendance: a case-crossover study. *BMJ: British Medical Journal*, 331(7514), 428. doi:10.1136/bmj.38537.397512.55
- Oviatt, S., Cohen, P., Wu, L., Duncan, L., Suhm, B., Bers, J., & ... Ferro, D. (2000). *Designing the User Interface for Multimodal Speech and Pen-Based Gesture Applications:*

- State-of-the-Art Systems and Future Research Directions. *Human-Computer Interaction*, 15(4), 263-322
- Reeves, L. M., Lai, J., Larson, J. A., Oviatt, S., Balaji, T. S., Buisine, S., ... & Wang, Q. Y. (2004). Guidelines for multimodal user interface design. *Communications of the ACM*, 47(1), 57-59.
- Reimer, B., & Mehler, B. (2013). The Effects of a Production Level Voice-Command Interface on Driver Behavior: Summary Findings on Reported Workload, Physiology, Visual Attention, and Driving Performance.
- Sternberg, R. J. (2009). *Cognitive psychology*. Belmont, Calif. : Thomson/Wadsworth, 2009.
- Stutts, J. C., Reinfurt, D. W., Staplin, L., & Rodgman, E. A. (2001). The role of driver distraction in traffic crashes. Washington, DC: AAA Foundation for Traffic Safety.
- Young, M. S., & Stanton, N. A. (2002). Malleable attentional resources theory: a new explanation for the effects of mental underload on performance. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 44(3), 365-375.
- Wärnestål, P., & Kronlid, F. (2014, February). Towards a user experience design framework for adaptive spoken dialogue in automotive contexts. In *Proceedings of the 19th international conference on Intelligent User Interfaces*(pp. 305-310). ACM.

# Using Tactile Detection Response Task for evaluating in-vehicle systems in trucks

Sofia Lindvall

4 Mars 2016

## Contents

<b>1</b>	<b>Introduction</b>	<b>3</b>
<b>2</b>	<b>The Tactile Detection Response task</b>	<b>3</b>
<b>3</b>	<b>ISO-standard 17488</b>	<b>4</b>
3.1	Background . . . . .	5
3.1.1	The multiple resource model . . . . .	5
3.1.2	Interference . . . . .	7
3.1.3	Strategies to handle demands . . . . .	7
3.1.4	Factors affecting DRT performance . . . . .	8
3.2	Method . . . . .	8
3.2.1	Participants . . . . .	8
3.2.2	Experimental set-ups . . . . .	9
3.2.3	Stimuli and response . . . . .	10
3.3	Procedure . . . . .	11
3.3.1	Instructions . . . . .	11
3.3.2	Training . . . . .	12
3.4	Data collection and analysis . . . . .	13
3.4.1	Performance measures . . . . .	13
3.4.2	Data analysis . . . . .	14
3.4.3	Data quality . . . . .	14
3.5	The ISO Coordinated studies . . . . .	15
3.5.1	Method . . . . .	16



3.5.2	Results . . . . .	17
3.5.3	Benefits and limitations with DRT . . . . .	20
<b>4</b>	<b>Previous studies</b>	<b>21</b>
4.1	National Highway Traffic Safety Administration (NHTSA) . . . . .	22
4.2	Tactile detection task as a real time cognitive workload measure. Cyriel Diels (2011) . . . . .	25
4.3	Evaluation of the tactile detection response task in a laboratory test using a surrogate driving set-up Roland Schindhelm, Eike Schmidt (2015) . . . . .	26
4.4	Comparison of static and driving simulator venues for the tactile detection response task. Johan Engström, Pontus Larsson and Christian Larsson (2013) . . . . .	27
4.5	The smartphone and the driver’s cognitive workload: A comparison of Apple, Google, and Microsoft’s intelligent personal assistants. Strayer, D. L., Cooper, J. M., Turrill, J. M., Coleman, J. R., & Hopman, R. J (2015) . . . . .	28
<b>5</b>	<b>The RIVER study</b>	<b>29</b>
5.1	Method . . . . .	29
5.2	Procedure . . . . .	30
5.3	Analysis method . . . . .	31
5.4	Lessons and recommendations . . . . .	31
5.5	Conclusion . . . . .	33

# 1 Introduction

In-vehicle systems in trucks have to meet the needs of the truck drivers as well as being safe. A user survey performed in the scope of the RIVER project showed that truck drivers have to perform several task while driving in order to do their job in a timely manner. As safety is a main priority when introducing in-vehicle system in trucks, ways of measuring the cognitive load the execution of these tasks requires are needed.

Detection response task (DRT) is a task that can be used for measuring how drivers' attention is affected by the demand of a secondary task (ISO 17488, 2015). The international Organization for Standardization (ISO) is currently working with developing a standard for the use of DRT to measure effects of cognitive load on attention for secondary tasks involving interaction with visual-manual, voice based or haptic interfaces. The National Highway Traffic Safety Administration (NHTSA) is developing voluntary guidelines for vehicle safety by discouraging excessive distractions such as in-vehicle and portable electronic devices (Ranney et al, 2014).

With recommendations from the studies mentioned above, a study design was developed and tested in a pilot study. This report will present information about the DRT as well as previous studies, the procedure of the pilot study and what lessons and recommendations that evolved from it.

## 2 The Tactile Detection Response task

ISO 17488 (2015) created a framework for understanding driver inattention, where they divided attention tasks into three levels of activity:

1. Sensory/actuator resources (lowest level, including the hands, the feet and eyes used for sensing the environment and create actions)
2. Perceptual/motor resources (brain structures that controls perceptual/motor resources, for example visual perception and manual tracking)
3. Executive control (higher level cognitive operations such as decision making, problem solving etc.)

According to ISO 17488 (2015) visual-manual tasks require all levels of resources, but mostly the first two levels. Many contemporary tasks can be

steered with voice interaction which can decrease distraction effects. However, according to Ranney et al. (2014) these task will mostly focus on the sensory/actuator and perceptual/motor resources and DRT has been developed in order to assess the effects of level three.

The DRT task is carried out by repeatedly presenting simple targets and recording the driver's response time. There are different types of DRTs available which differs in regard to what stimulus they present. The different variants of DRT mentioned by NHTSA (Ranney et al., 2014) and ISO 17488 (2015) are: *head-mounted* (HDRT), *tactile* (TDRT) and *remote* (RDRT). With HDTR the participant wears a fixture with a single LED above the left eye, which presents visual stimuli to the participant. An advantage with HDRT is according to ISO 17488 (2015) that the target remains in the same position relative to the driver's eye position, which eliminates the variability between the target and the head position.

With TDRT, which is the selected variant in the RIVER study, the driver has an electrical vibrator taped on the left shoulder. This variant has the same advantage as HDRT regarding the variability in position which is eliminated. Another advantage is that TDRT is not giving a visual stimulus. This eliminates conflict between detection of a visual target and the visual demand of driving, which according to Engström (2010) could mean that TDRT is the purest variant to measure attentional demand. The method has one known disadvantage: the use of vibration as a stimulus is not studied as much as the other stimuli types meaning that the effects of using it are not as well-known as for the other stimulus.

Finally, the RDRT, in contrast to the other two versions, consists of a LED placed in a fixed location in the center of the driver's view. This means that the target location isn't fixed relative to the driver, which might allow for lower sensitivity for detecting differences relative to TDRT and HDRT during visual-manual tasks where the driver looks away from the road.

### 3 ISO-standard 17488

This section is a summary of ISO-standard 17488 as described in Road vehicles Transport information and control systems Detection-Response Task (DRT) for assessing attentional effects of cognitive load in driving. The summary is focused on implications for TDRT, as this is the selected DRT variant in the study. The first part is a background covering relevant theory needed

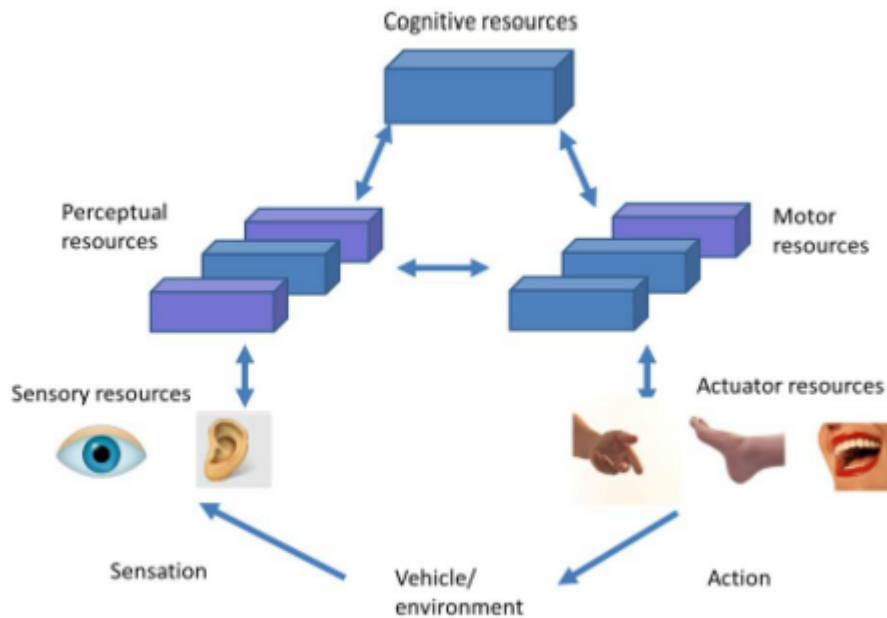
to understand the detection response task and what it measures. The next section covers recommendations regarding equipment, participants, training, procedure and the analysis of data. The third section presents important findings from an ISO-study conducted to validate the ISO-standard DRT procedure.

## **3.1 Background**

ISO-standard 17488 is a DRT used for assessing the effects of cognitive load on attention for secondary tasks involving interaction with visual-manual, voice based or haptic interfaces. According to ISO 17488 (2015) it is not possible to directly measure the cognitive load that a task demands. This can only be done by measuring how a task effects attention. The ISO-standard therefore has the main goal to state how to measure the effect of cognitive load on driver attention. Not to measure cognitive load directly.

### **3.1.1 The multiple resource model**

According to ISO 17488 (2015) all activities can be said to have two properties: activation (the degree of resources allocated to the particular activity) and selectivity (how much attention is allocated to the particular activity). The resources that can be allocated in activities can be divided to three levels: (1) Sensory/actuator resources, (2) perceptual/motor resources and (3) cognitive control. Cognitive control refers to higher level mental operations, for example decision making and problem solving. Intentional, top-down cognitive control requires mental effort and cognitive control is seen as a single resource with limited capacity. However, the demand gets reduced with practice, leading to increased automatized performance. The figure below is an illustration of this multiple resource model.



**Figure A.1 — Illustration of the 3-level resource model outlined in the text**

Figure 1: The multiple resource model (ISO 2015, 15)

Activities have different demands on resources. The ISO-standard (2015, 15) gives the following examples of loads for the different resources:

- Sensory/actuator load: the eyes to monitor the road ahead, or to view a display and demand for the hands to steer the vehicle or turn a knob on the radio.
- Perceptual/motor load: visual perceptual resources to detect a lead vehicle braking, or to perceive the content of a display, and manual motor resources to control braking or perform the radio knob turning action.
- Cognitive load: maintain items in working memory, deal with novel or inherently difficult tasks, or to overcome learned habits.

### 3.1.2 Interference

According to ISO 17488 (2015) activities and tasks can create interference between each other. The degree to which the task place overlapping demands on the resources needed to perform it, determines the interference between the activities and tasks. If two tasks are performed simultaneously the resources available might not be enough to support both tasks. This could lead to a degradation in performance of the task, depending on how the subject prioritizes between them. According to the multiple resource model task interference can occur parallel in the different resources. The detection response task is demanding in regards to all resources mentioned. Interference for the TDRT, specifically, is that sensory interference could occur if the tactile vibrations are hidden by other vibrations. If the hand used to respond also is used for a secondary task, actuator interference could occur. Motor resource interference could occur in all versions if the secondary task requires manual operations that are similar to the response. All versions are sensitive to interference between the attentional effects of cognitive interference when the secondary task demands cognitive control. ISO 17488 (2015) highlights that if the goal is to examine the effects of a task's cognitive load only, the DRT version to be chosen should be the one that minimizes the overlap of a DRT's sensory, perceptual, and response modalities with those of the secondary task that are being evaluated.

### 3.1.3 Strategies to handle demands

Drivers have different strategies to handle demands on different resources. It is important to note that they are not passive to task load. For example, a possible effect of this is that drivers' sacrifice performance on a task over another task, which may lead to the secondary task not being accurately reflected in the drivers' DRT performance scores. Drivers may have different strategies when performing the secondary tasks (resource allocation strategies may differ). Therefore, it is important to clearly instruct participants on how to prioritize between tasks and also to look at the performance of the secondary task and the driving task to check for effects of resource allocation strategies. Performance should be observed both when the task is performed alone, and in combination with other tasks.

ISO 17488 (2015) states that the TDRT could be preferable if a voice-controlled interface requires that the drivers sometimes glances away from

the roadway. The reason for this is that TDRT bypasses the visual modality and therefore has highest specificity for attentional effects of cognitive load. For the same reason TDRT is also preferable for visual tasks when the goal is to measure cognitive load. However, if the goal instead is to measure visual sensory and perceptual demand the RDRT is preferable. RDRT can also be preferable since it can be less intrusive than other DRT-variants. It may also be the preferable method, for the same reason, in case the sessions are relatively long.

### **3.1.4 Factors affecting DRT performance**

The following are factors that have been known to affect DRT performance (and which ISO 17488 (2015) recommends to always report):

- Vehicle type (passenger car, light truck, heavy truck or bus, etc.) and dynamics (in simulated driving the realism of the vehicle dynamics may affect DRT performance. Vehicle dynamics should be consistent in the study)
- Road type (urban, local road etc, road geometry)
- Road condition (dry, flat pavement is preferable; slippery conditions should be avoided)
- Traffic density (ideal traffic density is low to moderate. The driver should be free driving, meaning that the driver is not significantly affected by the presence of surrounding vehicles. If interfering traffic occurs, the test leader should avoid secondary tasks. If a secondary task is already started it should be interrupted and repeated later)
- Lighting conditions (daylight is preferable)
- Visibility (clear visibility is preferable; fog, heavy rain and snow should be avoided)

## **3.2 Method**

### **3.2.1 Participants**

According to ISO-standard 17488 (2015) participants should be drivers with licence and with a similar level of prior experience with the secondary task

under evaluation. The following relevant characteristics should be recorded for each subject:

- Driving experience (km driven in the last year)
- Similar device use experience
- Age
- Gender
- Experience with DRT

### 3.2.2 Experimental set-ups

ISO 17488 (2015) presents four different experimental set-ups that can be selected from and these are:

- Non-driving
- Surrogate driving
- Driving simulator
- On-road driving

*Non-driving set-up* means that a DRT and secondary task is performed at the same time without a driving component. In this case, attention is divided between the secondary task and the DRT. Baseline is DRT alone, with no secondary task or driving. For *surrogate driving*, DRT is performed at the same time as the secondary task and a primary task that is a surrogate for driving. For example, the driver may conduct a simple tracking task or watching a video of real-world driving. In this case, the baseline condition is DRT and the surrogate driving task, with no secondary task. For the *driving simulator set-up*, DRT, the secondary task and the driving task are performed at the same time. In this case, the baseline is DRT performed with the driving task without the secondary task. For the *on-road driving set up*, the DRT, secondary task and driving task are executed at the same time on a closed track or an open road. The baseline condition is DRT performed while driving.



### 3.2.3 Stimuli and response

ISO 17488 (2015) gives the following specifications as to how the stimuli set-up should be. The max stimuli duration is 1 sec. If the stimulus is present when the participant responds, the stimulus should be turned off at the same moment the participant responds. The stimulus cycle period refers to the time from the beginning of one stimulus to the beginning of the next stimulus. This should vary between a uniform distribution (a known number of outcomes equally likely to happen) of random values from 3 to 5 seconds.

For TDRT, which is the selected stimuli variant in the RIVER study, the stimulus is presented with a tactor, a small electrical vibrator, attached to the participant's left shoulder (given that the steering wheel is to the left, otherwise the tactor should be placed on the opposite shoulder). According to the ISO-standard 17488 (2015), the tactor can be placed on either shoulder in the non-driving set-up. The tactor can be attached using medical tape. The intensity of the vibrations of the tactor should be set to a level that takes the individual participant into consideration. Each participant must be able to feel vibrations saliently while at the same time not being in discomfort. It is also important to consider factors in the environment, such as vibrations from the road.



Figure 2: Placement of tactor (ISO 2014, 10)

The following table shows the default specifications recommended in ISO 17488 (2015):

Table 1: ISO default recommendations

Parameter	Value
Diameter	10mm
Weight	1.2g
Maximum speed	12 000rpm
Vibration amplitude	0.8g

The participant gets a micro-switch attached to the index finger, the middle finger or the thumb on the left hand (if the steering wheel is to the left). The participant clicks on this when being subjected to stimuli. If the set-up includes a driving task with a steering wheel the participant instead responds by pressing the micro-switch on the steering wheel. If the set-up does not involve driving and therefore has no steering wheel, the participant instead responds by pressing the switch against the desktop or the thumb.

### 3.3 Procedure

#### 3.3.1 Instructions

According to ISO 17488 (2015) the following instructions should be given to the participant at a minimum: overview of the purpose of the test, the expected duration of the test, and a presentation of the experiment procedure. Clarify that the purpose is to test how different tasks affect performance, and not to test the participant’s skills. Also, explain the secondary tasks along with general principles behind DRT and the primary task. The participant should be instructed to prioritize the primary task (driving or surrogate driving), but also, at a lower level, to prioritize the secondary task and the DRT. In the non-driving condition the participant should be instructed to do their best at performing the secondary task and DRT simultaneously. The following is an example of instructions from the ISO-standard (ISO 2015, 11) regarding task priority when the set-up involves driving:

*”Your main priority is to drive safely. Please remember to maintain your position within your given travel lane. The [LED/tactor] and the [secondary*

*task] task will both be active during the run. Please do your best to pay attention to both tasks but recall that your primary task is safe driving.”*

It is also recommended to explain to the participant that the data collection and analysis programs are designed so that they ignore responses when no stimuli is presented (ISO 17488, 2015). This means that a strategy of pressing the button without stimuli will not give better results.

### **3.3.2 Training**

All participants should receive training before the study starts. ISO 17488 (2015) recommends the following order: (1) the secondary tasks under evaluation, (2) the DRT and (3) the primary (driving or driving-like) task (if used in the study). Finally, the tasks should be practiced together. The ISO 17488 (2015) recommendations for each task type will now be presented individually.

#### **Secondary tasks**

A demonstration of the task is given by the experimenter. The participant gets a sufficient number of training trials (until stable performance and the participant feels comfortable). A guideline to use is the following: *“if 3 of the first 4 participants cannot successfully complete the practice task at least once in 5 trials, the interface design and training protocol should be reviewed”*. Record number of practice trials for each participant and task for post-test analysis. The data used in the practice trial should have the same complexity as the one used in the study, but the data should not be the same. An example could be that the practice trial and the study can have the same length of a street name when entering an address in a navigation task, but it should not be the exact same address. The experimenter should give the participant assistance and coaching if difficulties with the task occurs, so that the task is completed in an appropriate manner.

#### **DRT**

DRT training starts without any secondary task or driving. Training keeps going until the experimenter thinks that the participant respond to the stimulus in a stable manner and the participant feels comfortable. The experimenter should observe the participant to assure that the participant is trying to respond as quickly as possible and that the participant is not just clicking

the button without perceiving a stimulus. If these behaviors occur and the participant is not able to change that, the participant must be eliminated from the study.

### **Primary task**

The driving task without DRT or any secondary tasks. Training should continue until the participant reaches a stable performance and feels comfortable.

### **Multitask**

The last step in training is to perform all the tasks described above at the same time. If the study contains several secondary tasks, the multitasking should be trained for each secondary task. Continue training until the participant reaches stable performance and feels comfortable.

## **3.4 Data collection and analysis**

### **3.4.1 Performance measures**

The performance measures to be used are hit rate and response time (ISO 17488, 2015). A hit is a valid response to a DRT stimulus, which means a response initiated within 100 - 2500 ms from the stimulus onset and which is not preceded by an earlier response within the same interval.

An invalid response can be:

- Premature: before 100 ms from the onset
- Unrequested: initiated later than 2500ms after stimulus
- Onset: repeated, responses within 100-2500ms from onset that follows another response in the same interval

A missing response is when no response is present within the 100 - 2500 ms interval.

*Hit rate* refers to the number of hits divided by the total number of stimuli during a data segment (a continuous portion of data), excluding stimuli responded to prematurely.

*Response time (RT)*, which is the time from stimuli presentation to response, should be calculated for each hit. An RT value for a task or baseline is the mean response time for the relevant segment.

ISO 17488 (2015) states that a data segment must exceed 5 seconds in order to qualify for data analysis. For each task, at least five stimuli should be included in the analysis of each task. This can be done either by collecting multiple short data segments for the same task or by repeating the task within a single data segment until five stimuli have been presented.

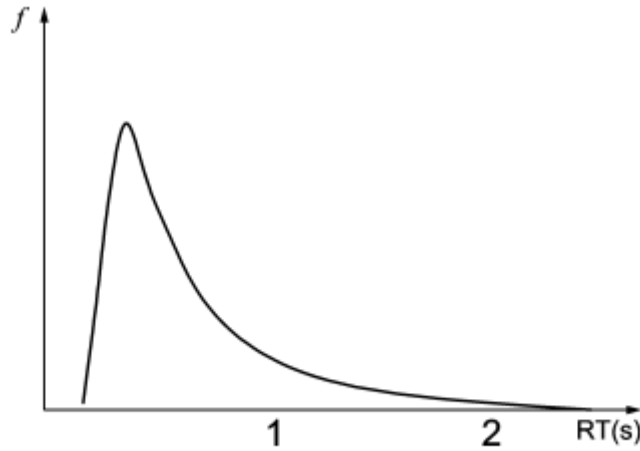
### **3.4.2 Data analysis**

ISO 17488 (2015) mentions several aspects to consider when analysing DRT data. Hit rates generally do not conform to normal distribution and the reason for this is for example a strong ceiling effect - most data points take a value of 1, meaning a 100% hit rate. In short data segments the hit rate can only take a limited number of discrete values. In a segment only containing three DRT stimuli, hit rate can only take four values, and the assumption of parametric statistical test (t-tests, ANOVA) is violated. In those conditions non-parametric tests are recommended instead. ISO also recommends caution to be taken when comparing tasks of short and long durations.

If hit rate and response time do not indicate an effect in the same direction, this could make the interpretation difficult. ISO exemplifies with when a secondary task under evaluation leads to reduction in hit rate, but no effect on response time. The reaction time should in that case be interpreted with caution.

### **3.4.3 Data quality**

ISO 17488 (2015) recommends to plot the frequency distribution of response times across participants into the form of a histogram. If the distribution deviates from the positive distribution below, the DRT measurement set-up should be controlled for possible technical issues.



**Figure 6 — Reference RT distribution for checking DRT data quality**

Figure 3: ISO (2015, 13)

Cheating strategies can be found by looking for the total number of responses for a subject and dividing it by the total number of stimuli for the subject. If the ratio exceeds 2, the data from the participant should be removed. ISO 17488 (2015) also mentions that video recording and monitoring during the session are good ways of identifying cheating strategies, such as repeatedly pressing the button without a present stimulus.

### 3.5 The ISO Coordinated studies

The objective of the ISO coordinated studies was to see if tests of secondary tasks, using the DRT methods in the standard procedure described above, produce reliable and valid results across different sites and set-ups. The tests had four tasks and a baseline condition, and used three DRT variants: TDRT, HDRT and RDRT.

These were the research questions:

1. To what extent do different RT results obtained with the DRT during non-driving, surrogate driving, driving simulator, and on-road set-ups?

2. To what extent are different RT results obtained with the TDRT, HDRT and RDRT?
3. Is the DRT RT sensitive and specific to the attentional effects of low vs. high levels of cognitive load for auditory-vocal and visual-manual tasks?

### 3.5.1 Method

Eight sites provided data for the analysis. Two tasks were used: *n-back task* and *surrogate reference task*. The n-back task is a delayed recall task where participants listen to and verbally repeat digits according to some rules, for example 0-back (recalls number that was said just before) or 1-back (recall the number that was said before the one just before). The n-back task gives auditory and vocal load, but no visual or manual load. Digits are spoken through a loudspeaker at a fixed interval for a 1- or 2-minute trial period.

Surrogate reference task (SuRT) is a search task that is self-paced. It requires visual and manual loads, but no auditory or verbal load. It is likely to also require cognitive control for example in regard to deciding the timing of when to press the right-hand button for the SuRT task and the left for DRT stimulus. The subjects scan a display for a target circle. The screen contains circles, and the target circle has a bolder stroke-width than the other circles. SuRT has two conditions: easy and hard. The circles are easier to visually discriminate with a larger difference in stroke-width in the easy version compared to the hard version. Selecting the target circle is done by pressing the left and right keypad buttons to move the grey outline bar to the target circle and press enter.

The task set-up for on-road, surrogate and simulator were triple-task conditions. Participants performed surrogate or driving, the DRT and the secondary task. The baseline was a dual-task condition of driving while performing the DRT. Non-driving was a dual-task condition with just DRT and a secondary task. The baseline for that condition was DRT alone with no secondary task or driving.

Trial times, repetitions and exposure times varied between sites. Trial times varied from 1 min to 4 min. According to ISO 17488 (2015), longer exposure times will present more DRT events and have more responses, which will lead to a reduction of uncertainty ranges in response times and the proportions of misses and hits. This will also give participants more experience

with the tasks which might reduce uncertainty in the estimates.

### **3.5.2 Results**

The following section covers a summary of the answers to each research question as described in ISO 17488 (2015).

#### **Question 1**

To what extent do different RT results obtained with the DRT during non-driving, surrogate driving, driving simulator, and on-road set-ups?

The relative pattern of response time result is the same for non-driving, surrogate driving, driving simulator and on-road set ups. The absolute response times are different for sites and set-ups, which is consistent with other research showing that response times never are the same between experiments conducted in the laboratory versus the road venue. It seems, however, that response times in a road test can be predicted by a laboratory test using the same task. This can be done by using linear regression which will correct for variations in absolute response time. For all driver performance metrics, DRT included, different set-ups and sites in driver performance research have relative validity, but poor absolute validity.

#### **Question 2**

To what extent are different RT results obtained with the TDRT, HDRT and RDRT?

The results show that faster absolute response times are acquired with the tactile version than with the head-mounted DRT version (HDRT). Faster absolute response times are also acquired with HDRT in comparison with the remote DRT version (RDRT). Relative response time results are the same regardless of DRT type.

A reason for this could be that there is an inherently faster response time for tactile stimuli than visual stimuli. Another explanation could be the fact that TDRT also gives a sound and that two sensory modalities will give faster response times than one sensory modality. This effect, if it exists, could be avoided by reducing the sound by putting the factor in a soundproof



enclosure, but this will also reduce the intensity of the factor which might produce slower response times than the visual versions of DRT.

### **Question 3**

Is the DRT response time sensitive and specific to the attentional effects of low vs. high levels of cognitive load for auditory-vocal and visual-manual tasks?

Findings show that DRT response time for audio-vocal tasks is sensitive, regardless of DRT type. However, it is not clear as to whether DRT response time captures the effect of cognitive load on attention for visual-manual tasks. Surrogate reference task (SuRT) showed that the response time for easy versus hard conditions do not differ in general during visual-manual tasks. Both the easy and hard condition had significantly longer response times than the 0-back task and slightly longer response times than the 1-back task. Explanations for this could be that the specificity of visual DRT versions could compromise the attentional effects on cognition to interfere with the visual modality. This does not explain why the results show that also TDRT found that easy and hard SuRT had longer response times than the 1-back task. A possible explanation to the lack of difference between the easy and hard conditions is that the participants are using pacing strategies. This means that participants slow down on the hard version to compensate for the difficulty, and perform with higher pace on the easy version. Another explanation could be that both SuRT and DRT need manual responses (pushing the button) which leads to a motor response conflict that gets solved with cognitive control. Easy SuRT has more button presses than hard, which leads to a stronger potential for motor conflict between easy SuRT and DRT. If this is a correct interpretation it means that it is best to be careful when interpreting data from response-intensive tasks.

### **Question 4**

To what extent are the results for hits/misses consistent with those for RTs?

A miss means that task load interfered with the ability to detect and respond to a stimulus. Of main interest are misses that are caused by the attentional effects of cognitive load, rather than those caused by visual load

from a secondary task.

Possible causes of misses:

- Not "seeing" a DRT stimulus: visual load leading subject to not seeing the stimulus. Measuring eyes off road and other glance metrics could capture the "glance portion" caused by a secondary task's visual load
- Not "shifting attention": Not shifting attention to the place where the stimuli appears

Different DRTs are sensitive to different causes of misses. TDRT, that will be used in the RIVER study, is sensitive to the attentional effects of cognitive load (if not tactile secondary tasks are used). TDRT response time is not affected much by visual load. However, it is expected to be affected by the effects of attention on cognitive load.

It is believed that the hit/miss rate or response time in a TDRT study will be affected in the same way as hit/miss and RT when the stimuli are visual or auditory. This is because the central attention effects to those stimuli are similar regardless of sensory modality. ISO (2015, 51) also writes that *"By studying the effect of a secondary task on the miss rate (or RT) to a tactile stimulus, we are assumed to be estimating the effect of a secondary task on the miss rate (or RT) to a visual stimulus, independently of any eye movements or blinks"*.

To study misses a miss rate analysis was made that showed that data was non-Gaussian, which means that conventional statistics (e.g. ANOVA) applied to hit/miss rate data will not give a valid estimate of the population. The same goes for mean, standard deviations and standard errors, etc. Instead, logistical regression analysis is more appropriate. ISO made a logistic regression analysis on the following premises:

- Only binary responses (a hit or a miss)
- No normality assumption is required
- Method was applied to individual hits and misses on a stimulus-level (the hits and misses for each individual subject are tabulated and analysed)

- Larger number of events than in a conventional analysis (depending on the point above)

Compared with ANOVA, there is a large increase in statistical power when using logistic analysis, and the ability to discriminate between task conditions is also much better.

The key finding from the miss analysis was that the probability of a miss for the visual DRTs exhibited a stronger sensitivity to the hard SuRT than to the easy SuRT. This is in contrast to the result for RT where none of the three versions of DRT distinguished between easy and hard conditions. Another finding was that the n-back task had a minor effect on DRT miss probability.

These are some of the benefits of performing a miss analysis:

- Measurement: Misses can give information on when the loading of a task is so high that it interferes to an extreme degree with attentional processes that slow responses for a stimulus down, and also leads to missed events
- Experimental control: misses can be used as a control variable for response time. For example, to ensure that participants are not using speed-accuracy tradeoff. It is important to examine hit/miss rate for different tasks. Two tasks can appear to have the same attentional effects from cognitive load if just RT is analyzed, but may have different attentional effects if hit/miss rate is analyzed
- Safety: DRT is supposed to address questions such as whether the driver is viewing the roadway (and is not tired) but still misses safety-critical events. For these questions, misses could be the main aspect making DRT safety relevant, more than RT

### 3.5.3 Benefits and limitations with DRT

The following are some of the benefits with DRT:

- DRT in a lab environment does not require extensive resources, yet gives the same relative results. It can therefore be used to improve human-vehicle interface design

- The ISO DRT procedure is repeatable across sites and stable in repeated tests
- Improving methods of analyzing data can be made by the methods of analysis developed to the study of missed events

The following are limitations with DRT:

- Not sufficient to prove if the attentional effects on cognitive load detected by the DRT have any role in creating crashes in real-world driving. According to ISO, a method able to detect that would require: a valid estimate of crash risk, valid measure of the attentional effects of cognitive load, a new way to connect estimate of crash risk and a measure of the attentional effects in naturalistic driving data
- Setting an absolute acceptance criterion will not work because the results varies between set-ups and sites. An acceptance criterion will lead to differences between sites, and meeting it or not will depend on where it was tested and with what DRT. An acceptance criterion should instead be based on a relative criterion, for example a ratio of a difference score from a baseline would provide relative ranking
- Cautions must be taken on the use of short exposure times or single trials. Exposure time is the product of trial time and the number of trials. Longer exposure time will enable the collection of more responses and stimulus
- Additional tasks are needed. Two levels of two task types were evaluated but the tasks should be more in numbers. The n-back task and SuRT are also non-ecological, meaning that they do not reflect real tasks
- The DRT methods that measures effect of cognitive load on attention should be used in addition to other driver performance metrics like eye glance to get the full picture of effects of secondary tasks while driving

## 4 Previous studies

Several studies have been conducted in the area of DRTs and what set-ups and variants that are preferable depending on different conditions. This section consist of an overview of some studies whose procedure and conclusions

are relevant in the planning and execution of a DRT study focusing specifically on TDRT.

## **4.1 National Highway Traffic Safety Administration (NHTSA)**

This section contains a summary of NHTSAs DRT study by Ranney et al. (2014).

### **Secondary tasks**

In the study performed by the National Highway Traffic Safety Administration (NHTSA) the n-back task was used to assess DRTs sensitivity for detecting differences in the attentional demands of secondary tasks. 0-back and 1-back were used.

Advantages with the n-back task is that levels of task difficulty and the attentional demand can be varied, and the fact that it is externally paced ensures a consistent level of task demand over time. Visual-manual tuning was also included. This was done to see if the DRT variants could provide comparable information when used to estimate secondary tasks with different interfaces.

### **Procedure**

In NHTSAs study two vehicles were used: one connected to a simulator that engaged the drivers in a car following task and in the other car the participants performed DRT and a secondary task without driving. They tested HDRT, TDRT and RDRT in both vehicles.

They had a repeated-measures within-subjects design, meaning that all participants performed all combinations of tasks. Number of participants were 48, 50% women and 50 % men. They were recruited in the following age ranges: 18 to 24, 25 to 39, 40 to 54 and 55 plus. Baseline in the driving simulator consisted of a task where the driver followed a car + DRT, and in the non-driving condition it was only DRT. The table below presents the procedure:

Table 1. Secondary Task by DRT by Test Venue Combinations

Trial	Secondary Task	DRT	Test Venue
1	0-back	Head-mounted (HDRT)	Static (Non-Driving)
2	1-back		
3	Radio tuning		
4	None (baseline)		
5	0-back	Tactile (TDRT)	
6	1-back		
7	Radio tuning		
8	None (baseline)		
9	0-back	Remote (RDRT)	
10	1-back		
11	Radio tuning		
12	None (baseline)		
13	0-back	Head-mounted (HDRT)	Driving Simulator
14	1-back		
15	Radio tuning		
16	None (baseline)		
17	0-back	Tactile (TDRT)	
18	1-back		
19	Radio tuning		
20	None (baseline)		
21	0-back	Remote (RDRT)	
22	1-back		
23	Radio tuning		
24	None (baseline)		

Figure 4: NHTSA procedure (Ranney et al 2014, 10)

The order in which the trials were presented was counterbalanced so that an equal number of participant had each factor in each possible position. Data collection interval was 3 minutes. One single stimulus was presented at temporal intervals selected randomly from times from 3 to 5 seconds. If the participant detected the stimuli it was immediately removed. The stimulus was present 1 second as a maximum.

### Data collection and analysis

Data collected were target activation time and target response time. Target activation time is time relative to beginning of trial that DRT target was

activated. Target response time is time relative to the beginning of the trial that DRT response was made. Both were collected in seconds. (Ranney et al 2014, 12).

### **Performance metrics**

The performance metrics were the hit rate and mean response time. Hit rate is proportions of targets correctly detected, which means number of targets correctly detected divided by the total number of targets presented in a data collection interval. Mean response time is the mean of all correctly detected targets during the same interval.

### **Data analysis**

The purpose of the analysis was to find the most sensitive and reliable DRT variants for each venue.

A comparison was made between the following secondary tasks:

- 0-back versus 1-back
- 0-back versus radio tuning
- 0-back versus baseline
- 1-back versus radio tuning

The data collection intervals on three minutes were first divided into 30 seconds and then 15 seconds. This was made in order to view patterns at different time intervals. For mean response time, the results showed that 15 second intervals is not enough and that 20 - 30 seconds is preferable. Mean hit rate value appeared to vary less over time compared to the response time means. Also, the difference among secondary task conditions was larger in the simulator venue than in the non-driving venue.

Ranney et al. (2014) defined the main questions in the study and provided answer to them. The main questions were:

1. Do differences exist among DRT conditions (HDRT, TDRT and RDRT) that would make one preferable for use in testing?

2. Do differences exist between test venues (Simulator, Non-Driving) that would make one preferable for use in testing?
3. Do differences exist among data collection intervals of different durations?

The answer on the first question is that TDRT has the advantage of being better at detecting the most challenging of the targeted differences and has slightly better test-retest reliability. TDRT was most sensitive to the differences in 0-back versus baseline. However, in the non-driving venue all DRTs were able to detect the difference. For the second question regarding differences between test venues it was found that in the non-driving venue, hit rates differences were weaker due to consistently high rates, reflecting ceiling effects. This means that this measure was not particularly useful in the non-driving venue. However, for response time there was more sensitivity among conditions in the non-driving venue. So if hit rate is a necessary metric a non-driving test venue might not be a suitable choice of venue. On question three, the result showed that a 2 minute interval is preferable. It was shown that in the simulator venue, TDRT and a 2 minute interval provided best results.

A question Ranney et al. (2014) discussed in addition to the question above is that HDRT and RDRT have conflicts with visual stimuli and visual target. As this is the case, information processing can be delayed. Another source of delay is also variations in head position when a target is present, a participant can for example look away when a target is present. This issue is primary a concern for RDRT. It could be a smaller issue for HDRT but it is no issue at all for TDRT (as the target is tactile and attached on the body). The issue is not the case for audio-vocal task without visual demand.

## **4.2 Tactile detection task as a real time cognitive workload measure. Cyriel Diels (2011)**

Diels (2011) conducted a simulation study (motorway and urban roadway) where the subject performed a tactile detection task (TDT) and a cognitive demanding counting task (both count aloud and count silent). The baseline task was TDT without driving for 4 min. The experimental design was a two-way (2x3) within-subjects design. The independent variables were driving environment and secondary task. The dependent variable were workload



(measured by TDT and NASA-TLX). A stimulus was presented once every 3 to 5 s and it was active for 2 s or until a driver responded to it. The results were that TDT found the difference in cognitive demand between the two counting tasks. This indicates that the TDT is sensitive even to small secondary task variations. Another benefit with TDT is a relatively high temporal resolution (precision of a measurement with respect to time) compared to other workload measures. Subjective measurement relies on introspection, and takes time from the primary task. TDT can track changes in workload with a temporal resolution of 4 s.

The following table displays the experimental drives used in the study:

**Table 1. Overview of experimental drives.**

<b>Familiarisation drive</b>	5–10 min (TDT; TDT + drive; TDT + drive + count)							
<b>Baseline TDT (Static)</b>	Time (min)							
	1	2	3	4				
<b>Exp conditions (Dynamic)</b>	1*	2	3*	4	5*	6	7*	8
M'way TDT								
M'way TDT + Count		Aloud		Silent		Aloud		Silent
Urban TDT								
Urban TDT + Count		Aloud		Silent		Aloud		Silent

\*Note that for the analysis, data collected in grey cells were not used for analysis.

Figure 5: Experimental drives in Diels (2011, 186)

### 4.3 Evaluation of the tactile detection response task in a laboratory test using a surrogate driving set-up Roland Schindhelm, Eike Schmidt (2015)

Schindhelm and Schmidt (2015) confirmed in their study that TDRT should not be used for task scenarios with a strong motor demand. They conducted a within-subject design with with four primary task variations (visual tracking with an easy or hard test track or auditory tracking with an easy or hard track), secondary task (SuRT, n-back, no task, and difficulty: hard and easy) and TDRT (with our without) as independent factors. The dependent factors were TDRT hit rate and response time.

They found that TDRT is intrusive on primary and secondary task performance. The analysis displayed that tracking task performance decreased. The reason for this could be that TDRT and the tracking task are manually operated which could lead to interference between the tasks. It also seemed that mental demands of TDRT intruded on primary task performance. This was seen especially when the cognitive demand of the task scenario was high, for example with the n-back task. A recommendation was that further studies should confirm sensitivity to secondary tasks and study intrusion on task performance more extensively.

TDRT could not differ between easy and hard SuRT. It was proposed that easy and hard do not differ that much in total visual-manual load, because manipulation of visual-manual load is able to be self paced because of the SuRT response frequency. Screens for SuRT occurred more often for easy SuRT than for hard SuRT leading to more manual responses for easy than for hard condition. Impairing effects of motor interference were more obvious in the easy than in the hard condition. The authors recommended to address this by avoiding this difference in motor demands by having tasks where self-regulation for manual workload is not possible for the participant.

#### **4.4 Comparison of static and driving simulator venues for the tactile detection response task. Johan Engström, Pontus Larsson and Christian Larsson (2013)**

Engström, Larsson and Larsson (2013) investigated static tactile DRT (no driving component) and found a preliminary result exhibiting that static TDRT is a valid low-cost alternative for driver interface evaluation when the interface is without excessive motor demands (such as voice control). In comparison, the more usual TDRT in a driving simulator venue is more demanding regarding costs and venue. However, they highlighted the need for more research, as for example the driving task in the study was relatively non-demanding and secondary tasks could have other effects in different driving conditions. To further test the validity of the method they suggested future studies with more naturalistic secondary tasks and driving situations.

The tasks used in the study where n-back, SuRT and a task involving Apple's speech recognition application Siri. All secondary tasks were performed for one minute and there were 30 seconds between tasks. Baseline data with

no secondary task was collected for one minute. Repetition of the tasks and baseline was two times per venue. The independent variables were TDRT venue (static, driving simulator), secondary task (0-back, 1-back, SuRT Easy, SuRT Hard, Siri and baseline) and task repetition (first, second). Task repetition was added to check for learning effects. The independent factors were varied within group and order of the secondary tasks and venue were counter-balanced between participants. The dependent variables were TDRT hit rate and response time. Venue order was included as a between-subject variable to check for training effects, as venue order determined how much training of the secondary task a participant got before the static venue condition.

#### **4.5 The smartphone and the driver’s cognitive workload: A comparison of Apple, Google, and Microsoft’s intelligent personal assistants. Strayer, D. L., Cooper, J. M., Turrill, J. M., Coleman, J. R., & Hopman, R. J (2015)**

Strayer et al. (2015) examined implications on cognitive load when using voice interaction while driving. They selected tasks that were free from visual demand and measured cognitive load with the Detection Response Task (DRT), the NASA TLX survey and video recordings. Their study found a significantly higher cognitive load when the driver was out on the road driving and at the same time executing the voice tasks, compared to when just driving. They did not carry out a comparison with a visual-manual counterpart, but added the highly cognitive demanding OSPAN task in an auditory variant. The results of cognitive workload between the OSPAN task and the voice tasks did not differ, meaning that the voice task imposed a high cognitive demand on the drivers. Another finding from the study was that the DRT data recorded exhibited that the cognitive load caused by interacting with the voice system lasted up to 18 second after the interaction. Strayer et al. (2015) explained this by proposing that the drivers need time to establish situational awareness.

## 5 The RIVER study

This section will describe the chosen method, the procedure and the lessons and recommendations that the pilot study gave.

### 5.1 Method

#### Participants

The study had seven participants (four male and three female). All were employees at Volvo trucks. They had all received drivers licence in order to work with development of the trucks. Therefore, they were not professional truck drivers. None of them had any earlier experience with the DRT.

#### Secondary tasks

The secondary tasks were to interact with the trucks in-vehicle system with voice as well as the visuo-manual counterpart. The tasks were the following:

- Call
- Note a reminder
- Play a track
- View the tacograph

#### Experimental set-up

An on-road set up was chosen in the study, which means that the participants drove in an open road. The TDRT, the secondary task and the driving task were executed at the same time. Baseline was the TDRT performed while driving.

The following are some factors that can affect the DRT, which ISO 17488 (2015) recommends to report. The vehicle that was used was a heavy truck which the participants drove in an urban area and on a motorway. The road conditions were mostly dry, but some of the participants drove parts of the experiment in rainy weather. The traffic density differed but was mostly moderate. It was daylight with clear visibility for all participants.

## Stimuli and response

Recommendations from ISO 17488 (2015) were followed to specify the set-up for stimuli and response for the DRT. The max stimuli duration was 1 sec and the stimulus was turned off at the same moment the participant responded. The stimulus cycle period refers to the time from the beginning of one stimuli to the beginning of the next stimuli. This varied between a uniform distribution (a known number of outcomes equally likely to happen) of random values from 3 to 5 seconds. The tactor was attached on the participants left shoulder. They also got a micro-switch attached to the index finger on the left hand to click on when they perceived the stimulus.

## 5.2 Procedure

Before the study, the participants received a letter with information about the purpose, the expected duration of the test, and a clarification that safety is the main priority. The participants were instructed to mainly prioritise the driving task and in second hand the secondary tasks and the TDRT. An instruction to respond as soon as the stimulus was perceived was also given.

The test lead differed between two personnel. The test lead assisted the participant with the DRT equipment and gave instruction throughout the test. He or she was also responsible for the computer software recording the DRT. Data was recorded in chunks for the entire baseline task, the speech task and the visual-manual task, including instructions and errors from the voice system.

In line with recommendations from ISO 17488 (2015) all participants received training before the study started. Because of time constraints, the training procedure in the study was modified from the one recommended in ISO. The participant first practiced the voice tasks, the visuo-manual tasks and the DRT while standing still. They then drove a test drive to feel comfortable behind the wheel. After that they drove away with the TDRT active and this was the baseline condition. The participant was asked to stop at a gas station to fill in a DALI questionnaire. When they drove off again they practiced the voice tasks while driving and short after they performed the voice tasks. They then stopped to fill in the DALI questionnaire again and drove off to practice the visuo-manual tasks. Following that, the visuo-manual tasks were performed. The participants were asked to drive back to the starting point to fill in the last DALI questionnaire.

### **5.3 Analysis method**

The preparation for data analysis of the DRT data was made with recommendations from ISO standard 17488 (2015). Hit rate and response time were the two performance measures that were used. The stimuli that were correctly detected and were in the ISO recommended range (100ms from onset and before 2500ms) were included in the analysis. Responses that were outside the scope were removed before data analysis. To cope with possible coping strategies from the participants, response times and hit rates were removed if the participant responded more than once to the same stimuli. The frequency distribution of response times across participants was plotted in the form of a histogram, which showed that the data were in the form of the distribution in Figure 3.

The mean response time and the mean hit rate was calculated for the baseline, speech and visual-manual task. The results were then presented in a bar graph. A repeated-measures ANOVA was used to compare the means from the three tasks.

### **5.4 Lessons and recommendations**

#### **Study design for comparing two interfaces**

The design of the study builds on the comparison of interacting with two different interfaces: a voice system and its visual-manual counterpart. The functions that should be evaluated in the voice system were decided first. The next step was to find the appropriate visual-manual tasks. A user survey performed in the beginning of the RIVER project identified in what ways the truck drivers in the study interacted with the truck while driving. These observations were used as guidance as to what visual-manual tasks to use. For example, the task of noting a reminder was decided to be in the form of writing information down on a piece of paper. This could have been done in other ways, for example by letting the participants use a smartphone and write a reminder in a note application. However, the observed participants in the user survey used paper and pen instead of their smartphones. When choosing tasks for the different interfaces, it is important that the tasks produce the same outcome. The procedure of the task will most likely differ, but the result from the task have to be the same so that an adequate comparison can be made.

## **Setting up the equipment**

When setting up the equipment, there are some aspects to consider. Whether the participant is left or right handed will affect the attachment of the micro-switch. It should be attached on the non-dominant hand, but it will affect the performance of the secondary task if it requires for example using the hand for clicking on buttons, as the case is with the visual-manual task. One way to solve this is to only include participants that are right handed.

The wire for the micro-switch could be in the way when the tasks requires bigger movements. Some participants complained that it was hard making turns because the wire was too short. Making sure that the wire is attached properly, loosely but still closely connected to the body, will avoid extra stress for the participants.

## **Instructions**

It is important to give the participants clear information in the same order. Especially the instruction to click as soon as a stimulus is perceived is important. Some participants might wait until the vibrations from the stimulus is over, and some might click as soon as possible. This difference will affect the collected data. The best way to solve the issue is to use the same test lead and have a detailed test guide. According to ISO 17488 (2015) it is important to tell the participants in what order the tasks should be prioritised. Some participant might (especially if driving on a closed track which is less safety critical) prioritise the TDRT in order to perform well. However, this will not reflect their performance in a real driving situation in an open road. Therefore all participant should be aware that the driving task is priority one, as safety is most important.

## **Training**

The participants in this pilot study had a different level of knowledge of the visual-manual tasks. The ISO standard 17488 (2015) recommends to have participants with the same level of knowledge of the secondary tasks, but this is not always an alternative as it can be difficult finding participants. When having participants with different levels of expertise in the tasks (both regarding the secondary tasks and the driving task), training will be even more important. In order to being able to make an adequate comparison between the baseline, the speech and the visual-manual task it is preferable

that each participant has the same level of stability in all tasks. However, training could be time consuming and during the pilot study the training procedure had to be cut down. This led to the test leads having to give assistance during the test, which should be avoided.

### **Task for training and test**

All participants started with the voice tasks in the pilot study. This should be counterbalanced in future studies to exclude training effects. Because of some practical issues, the participants did not always get the same tasks during training and the real test. This is a factor that could affect the result and future studies should take this into consideration.

### **Data analysis**

According to the ISO-standard 17488 (2015) recorded data from the TDRT for each task should include at least five data points. This could be done if the task is long enough, or by repeating the task several times. In this pilot study, the visual-manual tasks were all long enough to record more than five data points. However, the voice tasks were short, especially the task were the participants should make a call. Because of this, the analysis was made for DRT data for all voice tasks compared to DRT data for all visual-manual tasks. The DRT data recorded when the test lead gave instruction was also included in the analysis. The software for recording DRT data enabled the test lead to mark what task the data belonged to. But the tasks were so short that it was too time consuming to mark each individual task. In future studies, time points for when the different task starts should be noted, so that a more detailed analysis can be made for each task. The instruction should not be included in the analysis. Future studies could also put together more complex voice tasks that take longer time to perform.

## **5.5 Conclusion**

Safety is a main priority when introducing in-vehicle systems in trucks. A previous user survey in the RIVER project has shown that truck drivers have to execute some tasks while driving because of time constraints. This is something that need to be considered when designing in-vehicle systems



for trucks as the interfaces often expose cognitive load on the drivers. Ways of measuring the cognitive load truck drivers are exposed to are needed.

This report contains a description of a study design that was developed in order to evaluate a voice interface. The study design is a comparison of a voice interface with a visual-manual interface. The Tactile Detection Response Task (TDRT), which measures the attentional effects of cognitive load, was used to collect data. The result of the study is several lessons and recommendations that can be used in future studies.

## References

- Conti, A.S., Dlugosch, C., & Bengler, K. (2013) The Effect of Task Set Instruction on Detection Response Task Performance. In D. de Waard, K. Brookhuis, R. Wiczorek, F. di Nocera, R. Brouwer, P. Barham, C. Weikert, A. Kluge, W. Gerbino, & A. Toffetti (Eds.) (2014). Proceedings of the Human Factors and Ergonomics Society Europe Chapter 2013 Annual Conference.  
<http://www.hfes-europe.org/books/proceedings2013/Conti.pdf>
- Diels, C. (2011, May). Tactile detection task as a real time cognitive workload measure. In Contemporary Ergonomics and Human Factors 2011: Proceedings of the international conference on Ergonomics & Human Factors 2011, Stoke Rochford, Lincolnshire, 12-14 April 2011 (p. 183). CRC Press.
- Engström, J. (2010). The tactile detection task as a method for assessing drivers' cognitive load. Performance Metrics for Assessing Driver Distraction: The Quest for Improved Road Safety, 90-103.
- Engström, J., Larsson, P., & Larsson, C. (2013). Comparison of static and driving simulator venues for the tactile detection response task. In Proc. 7th International Driving Symposium on Human Factors in Driver Assessment, Training, and Vehicle Design (pp. 369-375).
- Ranney, T. A., Baldwin, G. H. S., Smith, L. A., Mazzae, E. N., & Pierce, R. S. (2014, November). Detection response task evaluation for driver distraction measurement application. (Report No. DOT HS 812 077). Washington, DC: National Highway Traffic Safety Administration.
- Road vehicles Transport information and control systems  
Detection-Response Task (DRT) for assessing attentional effects of cognitive load in driving. ISO/DIS 17488

Schindhelm, R., & Schmidt, E. (2015). Evaluation of the tactile detection response task in a laboratory test using a surrogate driving set-up. IET Intelligent Transport Systems.

# Subjective Usability Measurement for Speech Dialogue Systems In Trucks - A Methodology Study

Jonatan Andersson

January 18

## **Abstract**

This study aimed to find and modify subjective measurements for usability evaluation of speech dialogue systems (SDS) in trucks. While some already existing methods for evaluating of SDS's exists, we believe that our new and modified versions will get better data for usability improvement in this specific area. For instance, the subjective assessment of speech system interfaces (SASSI) was synthesized with the system usability scale (SUS) in order to be able to compare the SDS's with SUS's already validated and reliable database. In addition to this an interview based retrospective think-aloud protocol named Speech Think-Aloud Interview (STAI) was also created with the intention to generate rich qualitative data from the test participants. The methods were then tested on 7 participants with promising results.

# Contents

<b>1</b>	<b>Background</b>	<b>3</b>
<b>2</b>	<b>Introduction</b>	<b>3</b>
<b>3</b>	<b>Theory</b>	<b>3</b>
3.1	Differences between SDS:s and other UX-systems . . . . .	3
3.2	Learnability . . . . .	4
3.3	Questionnaires . . . . .	6
3.4	Subjective Assessment of Speech System Interfaces . . . . .	6
3.5	System Usability Scale (SUS) . . . . .	7
3.6	Think-Aloud Protocols . . . . .	9
<b>4</b>	<b>Speech Think-Aloud Interview (STAI)</b>	<b>11</b>
<b>5</b>	<b>SUSSI</b>	<b>13</b>
<b>6</b>	<b>Methodology</b>	<b>14</b>
6.1	Participants . . . . .	14
6.2	Tasks . . . . .	14
6.2.1	Baseline . . . . .	14
6.2.2	Visou-manual . . . . .	14
6.2.3	Speech Dialogue System . . . . .	14
6.2.4	Think-Aloud Session . . . . .	15
6.3	Data Analysis . . . . .	15
<b>7</b>	<b>Results</b>	<b>15</b>
<b>8</b>	<b>Discussion</b>	<b>15</b>
<b>9</b>	<b>Conclusion</b>	<b>16</b>

# 1 Background

This project was a collaboration between Linköpings University and a large truck manufacturer, partly financed by VINNOVA. The project aimed to research on how to promote safer roads with speech technology as its focus. While there has been many previous studies on speech dialogue systems (SDS) in vehicles before, the ambition with this particular paper was to find good methodology for evaluation of an SDS.

The system that is being evaluated in this study is based on a user needs-study that took part in 2015 which included observations and interviews both over phone and in person. A couple of features from that study were then implemented into a real truck which will be explained further in this paper.

# 2 Introduction

Speech dialogue systems (SDS) are becoming more common in society. As a result there is a need for validated and reliable ways to measure whether a SDS is good or not. This can be done by creating instruments that measure the usability of an SDS. It is however not an easy task as SDS's can be complex and take time to get comfortable with. A user could for instance know that a system has voice commands but still be clueless about what kind of interactivity with the system is allowed. Another potential issue may be that the user want to complete tasks that the system does not support by voice. Further, the context(s) where an SDS supposedly aims to be used is crucial to its usability as outer factors may distract or create errors for the user. The possible complexity and general issues with SDS's therefore calls for measuring instruments specifically designed to measure the usability of said systems. In addition to this there are different ways to measure usability and performance regardless of the product. The data that is being collected can for instance either be objective or subjective but may be analyzed individually or as a combination of the two. The research team evaluating an SDS has to decide which measuring instruments they believe will be the most suitable for their study as it is impractical to have them all due to time limitations and such.

This study is a pilot study for a larger clinic with the ambition to find good methodology to evaluate SDS with focus on usability and subjective measurements. While this study in particular will focus on the subjective measurements, a parallel study with the ambition to evaluate cognitive load will also place at the same time. The aim for this study is therefore to evaluate the methodology rather than what the methodology actually produces.

# 3 Theory

In this section the study's theory will be described and argued for.

## 3.1 Differences between SDS:s and other UX-systems

This section aims to describe fundamental differences between SDS's in contrast to other systems from a usability perspective. The sections purpose is to account for why it is necessary to have a specific subjective measuring instrument for evaluating the usability of an SDS and why a universal tool like the System Usability Scale (SUS) (1) is not an optimal solution. Finally, cognitive load will be an important factor as a mean to work towards safer roads as requested by VINNOVA.

A vital component for any SDS is its so called voice engine. The purpose of the engine is to understand what the user is saying and give proper answers. It is also important that this process is not slow as it will frustrate the user and make the system feel awkward to use. As a result of this speed is a crucial factor to account for when evaluating an SDS. Speed is however not a factor that makes SDS something special within the UX-practice. Palmer (2) for instance claims that any system with a response time greater than 200 ms will feel slow and frustrating to use.

Further, accents may be a serious issue when building a voice engine. The reason is that words will be pronounced in different ways as a result of cultural factors, which makes it is hard to create an engine that will be viable for the majority of its user. As English is the most commonly used language internationally it is required that you design for a great number of possible accents and pronunciations - not only those of native tongue. As a result of this it might be important to recruit test participants with different accents in order to truly test how powerful a voice engine is. The system will therefore act differently as a result of the user's voice which is something unique as it is something natural that cannot be changed. In order words a user for instance using a website may learn from errors while a user talking to a system may be punished for the way he or she is speaking which is something important to keep in mind.

For this particular study this is an important aspect in the evaluation of the system as a whole but the engine itself will not be calibrated.

One issue that designers struggle with when building an SDS is how to let the user know what voice commands are accepted by the system. It is therefore important that the user feels comfortable with the interaction between human and computer. As the user cannot reread information like one usually can when for instance browsing a web page, it is important that the information from the SDS is clear. As a driver for instance easily can be distracted on the road the option to have information repeated should be available.

As this study however aimed to evaluate a brand new SDS using subjective measurements, focus is on learnability which will be explained further in the next section.

### 3.2 Learnability

When designing a complex system it is vital to give the novice users tools to help them become experts. The system might be hard to use the first few tries, but as long as the user is satisfied and is making progress this should not be an issue.

When designing products in general, the usability aspect is often referred to. Nielsen (3; 4) defines usability as "a quality attribute that assesses how easy user interfaces are to use". Learnability together with efficiency, memorability, errors and satisfaction are the five quality components of usability. He defines the quality components as following:

- Learnability: How easy is it for users to accomplish basic tasks the first time they encounter the design?
- Efficiency: Once users have learned the design, how quickly can they perform tasks?
- Memorability: When users return to the design after a period of not using it, how easily can they reestablish proficiency?
- Errors: How many errors do users make, how severe are these errors, and how easily can they recover from the errors?

- Satisfaction: How pleasant is it to use the design?

Nielsen's definition of learnability however assumes that the user will "learn the design", but he does not propose a way to empirically check whether or not the user has done so. This might be a non factor when evaluating a small user interface as the system is easy to navigate through. It is however problematic when evaluating a large system as the user perhaps will not explore all its functions but still find it to be an effective and satisfying tool. The reason behind this is yet again unclear requirements as to when a user knows the system.

In contrast, Lehto (5) suggests that learnability in general terms could be defined as "a characteristic where performance improves with experience. As tasks are repeated, elements of the task are better remembered, prompts are more clearly distinguished, skills are sharpened, transitions between successive tasks are smoothed, eye-hand coordination is more tightly coupled, and relationships between task elements are discovered. The aggregation of these effects results in faster performance times, fewer errors, less effort, and more satisfied users".

This definition assumes that regardless of the task, there is no limit for improvement or skill. Further, as experience improves so will learnability as they are tightly coupled. This means that a system can never be learned in the definition that it is binary - you either know the system or you do not. As such, this definition is better suited for evaluation of a large and complex system with the assumption that not a single user will have complete knowledge of it. Listed below are some other definitions of learnability that has been used in both academia and practice:

- Nielsen (3): Novice user's experience on the initial part of the learning curve.
- Dix (6): Ease at which new users can begin effective interaction and achieve maximal performance
- Santos and Badre (7): Measure of the effort required for a typical user to be able to perform a set of tasks using an interactive system with a predefined level of proficiency.
- Hart and Steveland (8): The speed and ease with which users feel that they have been able to use the product or as the ability to learn how to use new features when necessary.
- Bevan and Macleod's (9): A measure of comparison the quality of use for users over time.
- Butler (10): Initial user performance based on self instruction and [allowing] experienced users to select an alternate model that involved fewer screens or keystrokes.
- Kirakowski and Claridge (11): Within the web context is the degree to which users feel able to manage the product's basic functions during its first use.

While the latter four are rather straightforward, the academic seems to disagree on the definition of learnability. Regardless of definition, learnability may be considered as the most vital usability quality as it is the user's first experience with the system. Further, while the other quality components of course are important one could argue that it is impossible to have a good level of learnability if the other components are bad. For instance, if the learnability is good (i.e. it is easy to learn how to use the system) one can assume that it is fairly easy to recover from errors or remember what commands are accepted by the system. As such, this study will aim to evaluate and assess the

learnability of the SDS with the belief that the other quality components will also be touched upon if issues are found during testing.

### 3.3 Questionnaires

When measuring in general it is expected that the instrument gives the same result on different occasions. Hone (12; 13) suggests that there are certain qualities required by the measuring instrument in order to do this. Listed below are some fundamental characteristics of good measurement:

Reliability (the results should be stable across repeated administrations). Validity (the technique should measure what it is intended to measure). Sensitivity (the technique should be capable of measuring even small variations in what it is intended to measure). Freedom from contamination (the measure should not be influenced by variables that are extraneous to the construct being measured).

In contrast to quantifiable metrics such as reaction times, the questionnaire approach is a subjective measuring tool that involves people's thoughts and feelings. Hone emphasizes that when the quality being measured is subjective the requirement for scientific rigour in the measuring tool is more difficult to achieve but still as important as with an objective. There are different reasons for this, such as the questionnaire's terminology and participants desire's to appear "normal" (social desirability effect). Further Hone suggests the discipline of psychometrics method's for developing valid and reliable measurement instruments as a way to achieve scientific rigour.

One question that is debatable in academics is how many values on a likert-scale a questionnaire should have (14; 15; 16). While some research do indeed claim that there will not be any differences between 5 and 7 points, other do find it. For this reason we want to use a 7 point scale as we could not find any literature that would advice us against it.

### 3.4 Subjective Assessment of Speech System Interfaces

Subjective assessment of speech system interfaces (SASSI) is a questionnaire with the purpose of accurately measuring a speech system's usability.

Hone (2014) lists the specific research objectives for SASSI as following: Valid, reliable, sensitive and free from contamination. Widely applicable to all styles of speech interface (for instance from command and control to natural language). Quickly and easily completed by naive and/or first time respondents. Quantifiable, to allow statistical comparison of multiple design alternatives, or benchmarking of a single product during development. Complete, capturing all important aspects of a user's experience with a speech system. SASSI consists of 7 questions and uses a likert scale from 1 to 7. The current state of SASSI shows promise, but is not yet a fully validated method for measuring usability of a system using speech in its interface.

What does SASSI actually tell about a system? The easiest answer is probably that it is possible to use the questionnaire in a summative way to compare two systems and then know which one of them is the best by comparing the numbers. A more advanced (and perhaps interesting) way to use SASSI in a more formative way is to analyze which sections a system will perform best



within. This can tell a designer where there is a need for further development and may give answers to questions about why user's are not giving the system higher ratings.

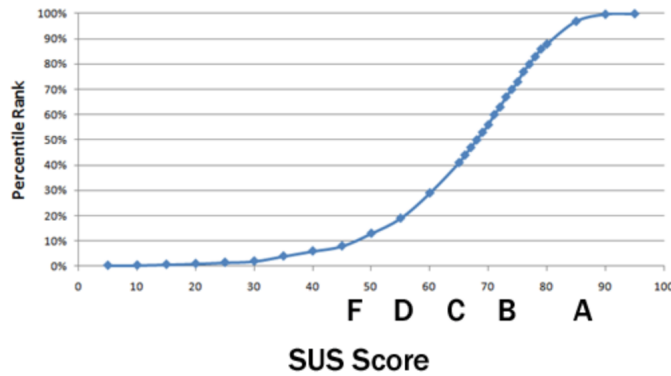
SASSI has previously been used for evaluating an SDS in vehicles. Hoffman et al. (17) for instance used SASSI together with the DALI questionnaire which derives from NASA TLX (18; 19) in order to assess both an SDS's usability and distraction. DALI (Driver Activity Load Index) is a questionnaire designed to measure subjective cognitive load when driving in comparison to quantitative data-collecting methods such as Driver Response Task (DRT).

There are however some problems with SASSI as the method is not fully validated yet. One potential issue is that the correlations between reality and the questionnaire's questions may be too weak. Some prefers to instead use the PARADISE-framework created by Walker et al. in 1997 (20) which combines subjective data with quantitative metrics such as task success. However, Hone and Graham argued that the PARADISE-framework's items chosen for user-satisfaction were not well-conducted or empirically based. Further, they do not agree with PARADISE way of summing all the test participants scores (21; 13). Indeed, summing all participants scores into one would make it impossible to find differences between users which may be an important factor. While combining quantitative metrics with subjective ones do indeed seem to be interesting, we believe that it may be better to only use subjective measurement for user satisfaction and instead aim to use metrics such as time on task, task success, logs and such to assess quantitative data.

Finally, another issue that there is no available database of reported SASSI results as the creators have not publicly released any data. This means that it is not possible to compare the SASSI results to a database, making it hard to know how good a system would do on the market. To counter this issue, we suggest that SASSI should be synthesized with the System Usability Scale (SUS) - questionnaire.

### 3.5 System Usability Scale (SUS)

The system usability scale (SUS) was initially designed to give usability practitioners a tool to quickly and easily assess the usability of a given system or product (1; 1; 22). The result was a questionnaire which nowadays has 10 items that accepts answers from a likert-scale, normally 1-5. A SUS questionnaire has a score between 1-100 which will tell how well a user appreciates a system's usability (23; 24). Despite only having 10 questions, research has shown that is one of not the most reliable questionnaire for assessing usability (22). Further, one of its strengths is that it is not limited to a specific product domain or area of use which makes it adaptable to any user-product relationship (25; 26). Since its creation the questionnaire has been validated and comes with a large database along with different ranking segments. These rankings will be listed below:



As the diagram tells us, anything below 51 would be considered an F or "fail". This rank would simply tell us that the system will struggle too much in practice and should therefore not be launched before its errors are fixed. Further, 68 seems to be the median level of all SUS-scores. As such, anything under 68 is below average and naturally anything over 68 is above average. Sauro also states that a SUS of 74 has higher perceived usability than 74 percent of all products tested. This score would also fall into the B-grade interval. To get the highest grade however, you would need a score of at least 80.3. This would not only land your system into the grade A-intervall, but is also believed to be at the level where the user will recommend the system to a friend.

Brooke J (1; 27) defines the questionnaire as "quick and dirty" as it will quickly give a brief overview of well the system is performing. The reason why he refers to it as "dirty" is because it is "fairly quick and dirty to administer". A perhaps deeper explanation as to how SUS may be dirty is because the shallow data it produces. To explain this further, consider that a system is being evaluated with the intention to improve it through usability testing. The SUS-questionnaire will quickly tell us whether or not the test participants finds the system user-friendly or not, but it will not tell us how to improve it. Instead it might at the very best tell us where the problem lays, but it will not get more specific than that. Instead, the SUS-questionnaire should be treated as a complement to richer methods such as the think-aloud protocol or metric-driven usability testing such as time on task, task completion etc.

Sauro (28) however agrees that SUS is by all means quick, but definitely not dirty. He emphasizes that SUS has data from over 5000 users and has been used in 500 different studies. Further, he suggests that its versatility, brevity and wide-usage means that despite inevitable changes in technology it is still being used. In addition he believes that SUS will still be around in 25 years due to its ability to be adapted to different areas of use.

Even though both Brooke and Sauro seem to have different opinions as to what "dirty" in this case refers to, I believe both have valid points. SUS is dirty as it takes little effort to administer, but it does not give dirty data. The problem however is that the data is not rich, which Sauro agrees on by simply admitting that SUS will not shed much light on why users are responding the way they are.

Why do we then use SUS instead of other perhaps richer methods? As SUS only has 10 questions, it does not take a lot of time to complete. Further, all the questions are on a likert scale 1-5 so the test participant does not have

to formulate answers. But then again, only because something does not take a lot of effort to complete does not mean that we should use it. Instead, the arguable most vital component of the SUS-questionnaire is simply that it will make it possible for us to compare our system's usability with others. This is especially important when evaluating new technology that uses new design patterns and interaction that previously has never been touched or researched. As SUS is both validated and has a large database, we can quickly tell if our new system is being appreciated by the user. It is also possible to see whether or not we are improving the system between iterations. In addition to this we can also more easily sell our ideas and system to product owners or other corporate organs as SUS has different grades depending on the score. While some research has found issues, such as that usability and learnability seem to be independent factors (29; 30), most do agree that it is indeed a quick way to get a good overview.

Academically, the number of users that is required for SUS to be reliable differs a lot. While some suggests that it takes at least 50 users to have appropriate data, Sauro for instance claims that the number of users is almost irrelevant and that even as low as 2 users are enough to get a good estimate. Nielsen however emphasizes that while some studies may be possible to get good results from with as low as 5 users, quantitative metrics require around 19 users in order to give good data. As the purpose of this study was to find differences between two systems, we think that using in-group design with 8 participants should be enough to see whether or not one system is preferred over another.

SUS has been around since 1996 and has since then been through some changes. One example is for instance changing the word cumbersome to awkward as the first word often confused the test participant (31). One greater change however has been proposed by Sauro and Lewis in 2011. They proposed that switching between negative and positive items in the questionnaire actually do not contribute to more valid and reliable data. Instead, it might only confuse the participant and make coding more difficult. He lists three reasons why mistakes might happen:

- Misinterpret: By switching between negative and positive questions users might respond differently as the reverse does not account for the difference.
- Mistake: Users may forget reversing their score and therefore accidentally agreeing with a negative statement when they meant to disagree.
- Miscoding: Researchers might forget to reverse the scales when scoring.

Further, Sauro's study showed that there were no significant differences between the original SUS and the all positive version. Thus, using an all positive SUS questionnaire appears to be reasonable as the feared question bias seems to be a non-factor.

### 3.6 Think-Aloud Protocols

The think-aloud (TA) protocol is one of the most commonly used tools when conducting usability tests as one cannot observe what a user is thinking (32; 33; 34; 35) while research has also shown that it will not greatly disturb the task (33). Nielsen (36; 37) defines the TA-protocol as:

“In a thinking aloud test, you ask test participants to use the system while continuously thinking out loud - that is, simply verbalizing their thoughts as they move through the user interface.”

Further, usability issues are found by encouraging the participant to verbally articulate what s/he is thinking or feeling when encountering a problem, and also how said problem can be solved. In addition to this the usability practitioners combine the TA-data with other metrics in order to identify and potentially solve as many issues as possible.

Many variations of the original TA protocol have been proposed with the two most common ones being the concurrent TA and the retrospective TA (38; 39). When using the concurrent TA, a test administrator will ask the participant to voice aloud thoughts, feelings and reasoning while completing one or more tasks using the system that is being evaluated. In contrast, the retrospective TA is used at the end of a testing session in order to collect the participant’s thinking and reasoning processes while they are still in the short-term memory of using a system. While both variations have their advantages, it is vital to choose method depending on what kind of system is being evaluated. For instance, research has shown that our gaze will be slightly disrupted when talking (40). If we then use the concurrent TA while also using eye-tracking equipment our data may be corrupted. Instead, the retrospective TA allows us to use the eye-tracking equipment without a disrupted gaze while still getting access to the user’s thoughts.

Regardless of what kind of TA is used it is crucial to carefully explain the methodology used in a study, which has been emphasized by many researchers. The main reason behind this is simply that it is close to impossible to replicate a study if the used TA is unknown. Additionally, the TA protocols used by practitioners may lead to inaccurate results.

Hawala et al. (39) states the problematic fact that there is a lack of empirical research studies on the TA protocol in usability testing. This has led to great variations between TA protocols which may lead to inaccurate usability results.

When writing this, there is still a lack of a validated TA protocol designed for speech dialogue systems (SDS). While a retrospective TA might be the obvious choice, one could argue that something has to be adjusted in order to optimally gather data from a user interacting with an SDS. In practice where time is often limited a video recording of a dialogue may not be time efficient or actually adding something as there may not be any visual feedback when using an SDS. Further, simply replaying a dialogue fragment for the participant may not give anything else than information about the terminology or a very specific function of the system. The essence of the problem when evaluating an SDS may therefore be to touch the system’s core and what strategy the user is supposed to use to be successful with the system. In order to counter the problem I propose a framework that a test leader may use to make the user session more efficient and to find the most fearsome usability issues with potential solutions.

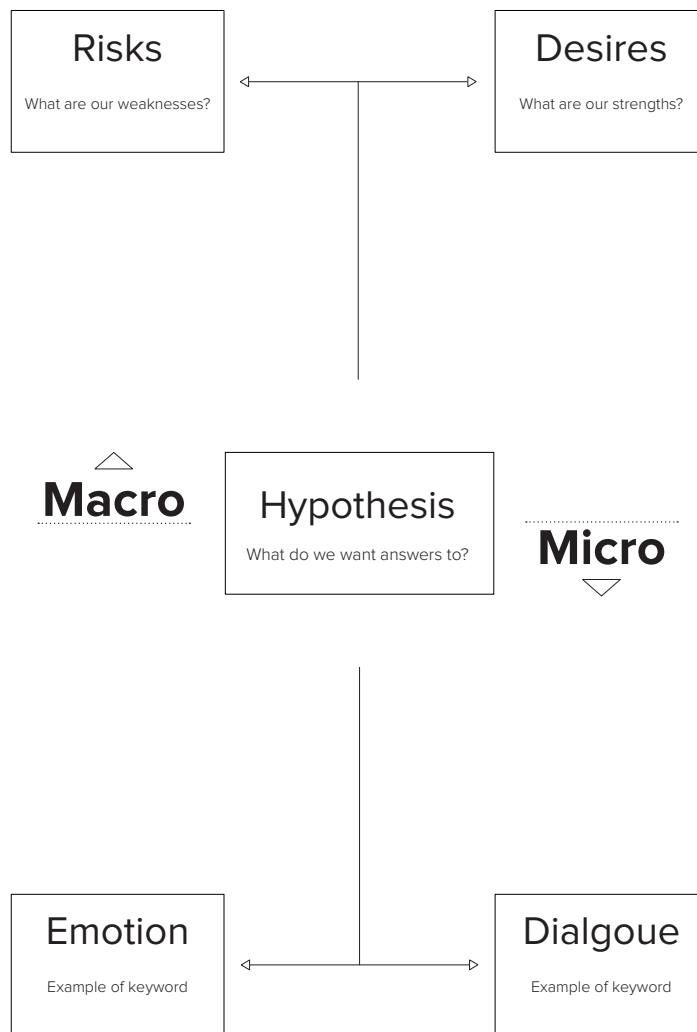
As an SDS often does not have any visual addition to the auditive output its navigation is limited to a space without visual feedback. As a result of this it might be difficult for the user to understand what input is accepted by the system and what kind of navigation that is allowed.

There are no detailed guidelines for how to use any sort of TAP. This have

lead to some problems in the academics as authors tend to miss explaining how they actually used their TAP. One reason why it may be important to explain how the TAP was used is to let the reader know how the test participant was instructed. As a TAP could significantly increase cognitive load it can negatively effect a test if there for instance is a time limit for the present task (41). Further, a moderator can actively encourage a participant by asking specific questions during the test but also remain silent throughout the whole session using the ordinary TAP. When using an RTAP questions may still be vague or specific and as mentioned earlier video will help the participant remember more easily.

## 4 Speech Think-Aloud Interview (STAI)

With the ambition to gather richer data while also enhancing RTAP, we suggest using a predefined guide when introducing the method for a participant. As the protocol will be created after instead of during interaction with a system we fear that the data collection will suffer if there are no set guidelines. Having a predefined guide would make the collected data more structured while also eliminating the risk of forgetting asking a question. When video recordings are available a guide may seem unnecessary as situations are specific and therefore hard to form questions for in advance. The session could however be divided into a macro/micro perspective, where the video accounts for the specific questions (micro) and the guide would ask more general questions (macro). This could potentially result in a broad and subjective evaluation of a system while also getting detailed information during specific scenarios. Another advantage is that all users will be able to reflect on the same areas in the macro-perspective which may show differences between them and their strategies. While this potentially may sound like a double edged sword as too much structure may limit the user's thoughts we encourage researchers to use a guide to extract as rich data from testing as possible - with efficiency. The idea behind the framework will be presented further below.



The idea behind this framework was to divide possible findings into a macro category and a micro category. The macro category would result in pre-defined questions of things that the design team may fear will be an issue. For instance, we might fear that our system is too complex and that our user's will get lost. This would be put under the Risks box. In contrast, we might have things that we want to confirm is good. For instance, we might want to be sure that our voice engines voice is good and satisfies the user. As such, we might simply ask the user if s/he likes it. Questions like these will end up under the desires box. The macro perspective will therefore generate questions that can be carefully adjusted to get as good answers as possible during the test session.

However, some things are simply impossible to foresee (and the reason why we do usability tests). These things will fall under the micro perspective and is based on what we observe during testing. It is therefore important to take notes or video record during the actual test session in order to help the user remember what was going on. For instance, if the user is struggling with a

voice feature we might take notes of what happened and when so we can ask questions about it. For instance, if we notice that the user is getting frustrated or confused, we will get a chance to assess his or her thoughts afterwards by trying to replay what happened. Both the macro and micro perspective aims give answers to our hypothesis or question. In our case for instance we want to know whether or not our SDS is good and secondly how we can improve it.

## 5 SUSSI

As the test sessions already was limited time-wise, a decision was made to synthesize SASSI and SUS, giving birth to a questionnaire with the working name SUSSI. The reason why SASSI was not considered sufficient enough to be used alone is that it is simply not validated nor comparable between other systems. The strength of SUS mentioned earlier is that it aims to assess usability without any given area which means that, in theory, the SUS will be able to compare an SDS to its validated database. Another reason why we believe that this is a reasonable choice is that most SUS questions have a good equivalent question in SASSI which means that the terminology will not differ much. Further, as both questionnaires aims to assess usability we do not believe that a lot will differ when we switch questions. In order to keep the questionnaire as short and efficient as possible, a decision was made to chose questions from the different SASSI segments as mentioned by Hone. The complete questionnaire is listed below where the 10 last questions originate from SUS.

1. The system makes few errors
2. I was able to recover easily from errors
3. I felt tense using the system
4. I felt calm using the system
5. A high level of concentration is required when using the system
6. The interaction with the system is frustrating
7. I sometimes wondered if I was using the right word
8. It is easy to lose track of where you are in an interaction with the system
9. The system responds too slowly
10. The interaction with the system is consistent
11. It is clear how to speak to the system
12. It is easy to learn to use the system
13. I would use this system
14. I felt in control of the interaction with the system
15. I felt confident using the system
16. The system is easy to use
17. I always knew what to say to the system
18. The system is simple
19. I found the various functions in the system were well integrated

## 6 Methodology

This section will describe the methodology used in this study.

### 6.1 Participants

7 employees from the truck manufacturing company were recruited for this study with the only requirement were that they had a truck license. As this study only aimed to evaluate the methodology, no demographical data was collected.

After getting into the vehicle, the participants were given the instructions for the test both in text and out loud. Before starting the truck, it was emphasized that the first priority during the test would be to drive safely and that they should fulfill their tasks when they felt comfortable to do so.

### 6.2 Tasks

A total number of 5 tasks were used in this study based on the previous user-needs study.

1. Make a call
2. Play a track
3. Remind me to..
4. Find out when the next break is
5. Check service light

Before driving off, the participants were instructed and shown how to complete the tasks both by using the SDS and visou-manually. They also got to feel how the DRT equipment would work in practice.

#### 6.2.1 Baseline

The bassline session consisted of driving from a town to a gas station approximately 15 minutes away. During bassline the participants only used DRT and as such no other tasks. After bassline the participants got to fill out a DALI-questionnaire as part of another study.

#### 6.2.2 Visou-manual

After bassline, the participants were instructed to complete the tasks mentioned earlier. They did however get coaching from the test leader and would therefore not have to remember the task completely. After completing the tasks the participant filled out the SUSSI and DALI questionnaire.

#### 6.2.3 Speech Dialogue System

As a final part of the driving session, the participants used the SDS to complete the tasks. After doing so, they filled out another SUSSI and DALI questionnaire. While interacting with the systems, additional data was collected by the note taker. The reason for this was to be able to discuss the SDS with the test participant afterwards as part of the TAP-segment.



#### 6.2.4 Think-Aloud Session

After arriving back at the starting point of the session, the think-aloud session began. Notes that were taken during the testing were highlighted in order to get insight from the participant.

### 6.3 Data Analysis

The questionnaires were analyzed with IBM SPSS Statistics version 23 while the data from the think-aloud session was analyzed from notes and sound recordings by coding themes.

## 7 Results

Significant differences were found between the visuo-manual (VM) and speech dialogue system (SDS) by analyzing the questionnaires. SDS had significantly better SUS and SASSI ratings than VM ( $p < 0,00$ ).

The data from the TA-session resulted in the following findings which all participants agreed to (7/7):

1. The voice was pleasant to listen to.
2. The system's feedback was bad. The participants did not always know when they were supposed to talk or when the system was thinking.
3. Some thought that the voice spoke too slow. This might however be a result of the bad feedback.
4. Sometimes they did not feel understood by the system. This resulted in uncertainty for what kind of input that would be accepted. One user for instance said "yes please" which the system did not understand because of his pronunciation. The user then thought that the system did not accept pleasantries such as "please" because of this failed communication. This would suggest that there are learnability issues with the system.
5. They thought that all features were good and relevant.
6. No one preferred to do a task visuo-manually instead of with the SDS.

## 8 Discussion

This study showed two promising approaches towards subjective evaluation of SDS's, SUSSI and STAI. While the methods did well in this study, the approach is however far from reliable or validated however. As this study was a pilot for a greater clinical session, we hope to get more insight and possible improvements of both methods. Further, more iterations of the system has to be done in order to see how well the STAI actually works and whether or not it is even close to the original RTAP. Further, even though STAI found usability issues, the current methodology of this study does leave room for improvement. Firstly, the first interaction with the system should not be guided. Instead the users should navigate and explore the system themselves instead of getting coached throughout the whole design. Secondly, the system is far

too small to let the user have a longer dialogue with the system. It is in other words too simple to get rich data from as the features do not allow complex input or such. This was sadly a letdown as we had hopes about assessing and evaluating learnability, but our study design probably did not help us in doing so. This was however a necessity as we also wanted to use DRT equipment for another paper. Finally, with this being said, it is important to once again emphasize that this was a pilot study with the ambition to find good methodology rather than interesting results from the actual methodology. For instance, the results found from the questionnaires tells us that it is better to use an SDS in comparison to a visou-manual system when you are supposed to focus on driving. This did not come as a surprise as the tasks probably were in the SDS's favor even before the test started. However, as this study did get financed on the terms that it would improve road safety it was a necessity to compare things that truck drivers actually do on the road today with things that they hopefully will do in a near future.

Finally, even though the actual results from this methodology did not give us anything really exciting, we do think that our thorough literature study has given us enough meat on our bones to at least give us reason to believe that there is something valuable here. We do however think that a lot of further work has do be done in order to produce something that can be used with good validity and reliability both in practice and academically.

## 9 Conclusion

A thorough literature study resulted in a synthesized questionnaire named SUSSI and a new approach to assessing people's thoughts which we call STAI. While the two methods did not produce any interesting findings, we still believe that they performed well enough to be considered using again with greater results. Further, a possible reason as to why the results were not more interesting may be that the system that was being evaluated was too simple and therefore did not generate richer data. Finally we believe that these tools have good potential but must be used and calibrated in order to make sure they live up to our standard.

## References

- [1] John Brooke. Sus-a quick and dirty usability scale. *Usability evaluation in industry*, 189(194):4–7, 1996.
- [2] Jonathan W Palmer. Web site usability, design, and performance metrics. *Information systems research*, 13(2):151–167, 2002.
- [3] Jakob Nielsen. *Usability engineering*. Elsevier, 1994.
- [4] Jakob Nielsen. *Usability 101: Introduction to Usability*, 2012 (Accessed January 18, 2016). <https://www.nngroup.com/articles/usability-101-introduction-to-usability/>.
- [5] Mark R Lehto and Steven J Landry. *Introduction to human factors and ergonomics for engineers*. Crc Press, 2012.
- [6] A Dix, J Finlay, and G Abowd. *R. beale: Human-computer interaction*, 1998.
- [7] Paulo J Santos and Albert Badre. Discount learnability evaluation. 1995.
- [8] Sandra G Hart and Lowell E Staveland. Development of nasa-tlx (task load index): Results of empirical and theoretical research. *Advances in psychology*, 52:139–183, 1988.
- [9] Nigel Bevan and Miles Macleod. Usability measurement in context. *Behaviour & information technology*, 13(1-2):132–145, 1994.
- [10] Keith A Butler. Connecting theory and practice: a case study of achieving usability goals. In *ACM SIGCHI Bulletin*, volume 16, pages 85–88. ACM, 1985.
- [11] Jurek Kirakowski, Nigel Claridge, and Richard Whitehand. Human centered measures of success in web site design. In *Proceedings of the Fourth Conference on Human Factors & the Web*, 1998.
- [12] Kate Hone. Usability measurement for speech systems: Sassi revisited.
- [13] Kate S Hone and Robert Graham. Subjective assessment of speech-system interface usability. In *INTERSPEECH*, pages 2083–2086, 2001.
- [14] John G Dawes. Do data characteristics change according to the number of scale points used? an experiment using 5 point, 7 point and 10 point scales. *International journal of market research*, 51(1), 2008.
- [15] Michael S Matell and Jacob Jacoby. Is there an optimal number of alternatives for likert scale items? study. *Educational and psychological measurement*, 31:657–674, 1971.
- [16] Kraig Finstad. Response interpolation and scale sensitivity: Evidence against 5-point scales. *Journal of Usability Studies*, 5(3):104–110, 2010.
- [17] Hansjörg Hofmann, Vanessa Tobisch, Ute Ehrlich, and André Berton. Evaluation of speech-based hmi concepts for information exchange tasks: A driving simulator study. *Computer Speech & Language*, 33(1):109–135, 2015.

- [18] Sandra G Hart. Nasa-task load index (nasa-tlx); 20 years later. In *Proceedings of the human factors and ergonomics society annual meeting*, volume 50, pages 904–908. Sage Publications, 2006.
- [19] Dagmar Kern, Angela Mahr, Sandro Castronovo, Albrecht Schmidt, and Christian Müller. Making use of drivers’ glances onto the screen for explicit gaze-based interaction. In *Proceedings of the 2nd International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, pages 110–116. ACM, 2010.
- [20] Marilyn A Walker, Diane J Litman, Candace A Kamm, and Alicia Abella. Paradise: A framework for evaluating spoken dialogue agents. In *Proceedings of the eighth conference on European chapter of the Association for Computational Linguistics*, pages 271–280. Association for Computational Linguistics, 1997.
- [21] Melita Hajdinjak and France Mihelič. The paradise evaluation framework: Issues and findings. *Computational Linguistics*, 32(2):263–272, 2006.
- [22] Thomas S Tullis and Jacqueline N Stetson. A comparison of questionnaires for assessing website usability. In *Usability Professional Association Conference*, pages 1–12, 2004.
- [23] Aaron Bangor, Philip Kortum, and James Miller. Determining what individual sus scores mean: Adding an adjective rating scale. *Journal of usability studies*, 4(3):114–123, 2009.
- [24] Aaron Bangor, Philip T Kortum, and James T Miller. An empirical evaluation of the system usability scale. *Intl. Journal of Human–Computer Interaction*, 24(6):574–594, 2008.
- [25] Panagiotis Zaharias and Angeliki Poylymenakou. Developing a usability evaluation method for e-learning applications: Beyond functional usability. *Intl. Journal of Human–Computer Interaction*, 25(1):75–98, 2009.
- [26] Philip T Kortum and Aaron Bangor. Usability ratings for everyday products measured with the system usability scale. *International Journal of Human-Computer Interaction*, 29(2):67–76, 2013.
- [27] John Brooke. Sus: a retrospective. *Journal of Usability Studies*, 8(2):29–40, 2013.
- [28] Jeff Sauro. Measuring usability with the system usability scale (sus), 2011 (Accessed January 18, 2016). <http://www.measuringu.com/sus.php>.
- [29] James R Lewis and Jeff Sauro. The factor structure of the system usability scale. In *Human Centered Design*, pages 94–103. Springer, 2009.
- [30] Simone Borsci, Stefano Federici, and Marco Lauriola. On the dimensionality of the system usability scale: a test of alternative measurement models. *Cognitive processing*, 10(3):193–197, 2009.
- [31] Kraig Finstad. The system usability scale and non-native english speakers. *Journal of usability studies*, 1(4):185–188, 2006.

- [32] K Anders Ericsson and Herbert A Simon. Verbal reports as data. *Psychological review*, 87(3):215, 1980.
- [33] Ronald P Leow and Kara Morgan-Short. To think aloud or not to think aloud: The issue of reactivity in sla research methodology. *Studies in second language acquisition*, 26(01):35–57, 2004.
- [34] Jakob Nielsen. *Thinking Aloud: The 1 Usability Tool*, 2012 (Accessed January 18, 2016). <https://www.nngroup.com/articles/thinking-aloud-the-1-usability-tool/>.
- [35] Jakob Nielsen. *When to Use Which User-Experience Research Methods*. <https://www.nngroup.com/articles/which-ux-research-methods/>.
- [36] Jakob Nielsen. *Demonstrate Thinking Aloud by Showing Users a Video*, 2014 (Accessed January 18, 2016). <https://www.nngroup.com/articles/thinking-aloud-demo-video>.
- [37] Janni Nielsen, Torkil Clemmensen, and Carsten Yssing. Getting access to what goes on in people’s heads?: reflections on the think-aloud technique. In *Proceedings of the second Nordic conference on Human-computer interaction*, pages 101–110. ACM, 2002.
- [38] Maaïke Van Den Haak, Menno De Jong, and Peter Jan Schellens. Retrospective vs. concurrent think-aloud protocols: testing the usability of an online library catalogue. *Behaviour & information technology*, 22(5):339–351, 2003.
- [39] Erica L Olmsted-Hawala, Elizabeth D Murphy, Sam Hawala, and Kathleen T Ashenfelter. Think-aloud protocols: a comparison of three think-aloud protocols for use in testing data-dissemination web sites for usability. In *Proceedings of the sigchi conference on human factors in computing systems*, pages 2381–2390. ACM, 2010.
- [40] Sanne Elling, Leo Lentz, and Menno de Jong. Retrospective think-aloud method: using eye movements as an extra cue for participants’ verbalizations. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 1161–1170. ACM, 2011.
- [41] Maarten W Van Someren, Yvonne F Barnard, Jacobijn AC Sandberg, et al. *The think aloud method: A practical guide to modelling cognitive processes*, volume 2. Academic Press London, 1994.

Reduced dIstraction via Voice intERaction –  
RIVER deliverable D2  
May 2015

Annika Silvervarg, Sofia Lindvall, Jonatan Andersson, Arne Jönsson

This report is part of the project RIVER - Reduced dIstraction via Voice intER-action. RIVER aims to explore the relationship between visual and cognitive load and voice interaction. The general objective of the project is to assess, develop and demonstrate new improved voice functionality with the goal to improve safety and efficiency and at the same time create a high user-acceptance.

In WP2 the focus is on the following questions, which are addressed in this report:

- How do truck drivers communicate today (e.g. what functions and with what tools – brought-in, build-in) and what speech interaction needs could they possibly foresee?
- What features in trucks would drivers prefer to control by voice, additional to the ones already controllable by voice?
- How can Volvo's voice control interaction be made more efficient? E.g. non-menu based system, more natural speech?
- How can apps and other nomadic device related features be safely integrated via voice control system run in the vehicle?

To answer the two first questions user needs studies including focus group interviews, observations of drivers and interview with drivers have been conducted. Two focus groups were conducted as a first step to get information about Volvo's perspective on voice control in trucks - the first focus group was held with personnel working with Human-Machine Interaction (HMI) and the second group with test drivers. After that, participant observation was conducted with truck drivers on one of their normal day's work out driving. Participant observation was conducted in the truck cab whilst on the road driving. A semi-structured interview was conducted during the driving session. The participant was asked about his or her goals and needs in their work, questions regarding attitudes towards voice interaction and earlier experience with voice interactions. Interviews were also conducted where the drivers were shown pictures and discussed three different themes/concepts. These more qualitative studies were followed up by a quantitative interview study with 70 truck drivers. The results from these are summarised in section 1 in the form of functionality available through speech desired by truck drivers. More details regarding these studies can be found in the reports "Concept Design for Voice Control in Trucks - What do drivers actually want?" by Jonatan Andersson and "Speech recognition technology in trucks: potential uses and implications for visual-manual distraction" by Sofia Lindvall.

To provide a theoretical base for answering the two last questions a survey of the latest technical development including recent results from research on design, development and evaluation of dialogue systems in cars and an overview of speech interfaces related to automotive on the market have been performed and presented in the report "Automotive Speech Interfaces - A State of the Art survey" by Annika Silvervarg and Arne Jönsson. The findings from the survey are summarised in the form of challenges and guidelines to consider during implementation of speech interfaces in cars in section 2.

# 1 What do drivers desire in speech interfaces?

Focus group interviews performed within the project with both developers at Volvo and test drivers in Hällered had the purpose of collecting a first set of speech functionality that were considered useful and could potentially increase the safety of driving trucks. The results are summarised in Table 1 and Table 2.

Table 1: Voice control functions suggested by Volvo developers

No.	Information to be read out loud
1.	Traffic information
2.	Vehicle information
3.	Status checks
4.	Coaching messages
5.	Warnings messages
Voice commands	
6.	Ask about break times and spots to stay at
7.	Handle orders and e-mails
8.	Connect to road trains
Desired effects	
9.	Control functions from other parts of the truck
10.	Being able to switch back to manual steering
11.	An adaptive system where the driver can map up own functions

The focus groups were followed up with participant observations. The recruited participants for the participant observation and interview sessions were truck drivers that represented different work segments defined by Volvo Group Trucks. The work segments were: city distribution, regional distribution, interregional distribution, demanding long haul, light construction, heavy construction and heavy transport. Eight truck drivers participated in the study. Five occasions consisted of participant observation and three were phone interviews. Difficulties with finding a long haul driver and a heavy transport driver led to phone interviews with two long haul drivers and one heavy transport driver. A participant observation was conducted with an additional driver from the city distribution segment.

The following main categories were identified during the analysis:

**Driver information** Three personas created in purpose of providing an understanding of the truck drivers, including working situation and attitudes.

**Distractions** Visual-manual distractions that the truck drivers were subjected to.

**Problems** Problems encountered by the drivers during their daily work.

**Suggested functionality** Voice control functionality explicitly suggested by the truck drivers.

Table 3 shows an overview of the distractions, problems, and suggestions and the type of voice commands they represent.



Table 2: Voice control functions suggested by test drivers at Hällered

No.	Information to be read out loud
1.	Text messages and e-mails
2.	Red warnings
3.	Rules and laws when crossing land borders
Assistance	
4.	Control functions from other parts of the trucks, for example temperature and lightening from the bed
5.	Traffic messages in other languages translates
6.	Traffic signs translated and read out loud
7.	Get information about resting times – when they need to stop, where and if it has available parking
8.	Ask the truck’s digital handbook questions and get answers read out loud
9.	Navigation system helping with planning routes based on the trucks weight
10.	Navigation system warning the driver if not reaching the destination without refuel or taking a break
Desired effects	
12.	Remove buttons for more space
13.	Focus on the road
14.	Save personal settings between trucks
15.	Activate voice control by pressing a button

The project also included interviews with the participants where they were where shown pictures on an iPad. The purpose of the pictures was to help the participant trigger ideas and memories by showing different scenarios relevant to the themes that had been found in the data from the focus group discussions. The technique is simply referred to as "triggers" and include any method that helps a person to imagine something using an object or a picture. For the phone interviews that were held, the scenarios had to be carefully described to compensate for the lack of actual trigger pictures. Other than that the procedure was not changed. After the scenarios had been presented, an open conversation was held where the driver could add anything that came to mind that had not already been covered by the interview guide.

The first concept was "I am sitting in a new truck. I cannot find the buttons that I am looking for, I am getting messages that I do not understand and I have paid a lot of money for functions that I do not even know how to use. What am I going to do?" and the resulting functionality desired to deal with this was:

1. Digitalize the car’s manual. By storing a digital manual in the cars interface one could search through it by using words or by speech and by doing so finding information about function, error messages etc. more efficiently. The manual enables the user to use the vehicle maximally.
2. Make the car’s interior smart. If the user would say "what does this button do?" and then press on one of the vehicles analog buttons, the user is going to get a short introduction of said button and get the choice to get further information.

Table 3: Observed distractions, problems and suggestions for voice control solutions

<b>Distraction (D)/ Problem (P)/ Suggestion (S)</b>	<b>Frequency</b>	<b>Voice control solution</b>
Text messages and phone calls (D)	8	Connect phone to infotainment unit and enable voice interaction
Navigation (smartphone applications and GPS) (D)	4	Connect phone to infotainment unit, Navigator with speech recognition
Writing information down (D)	3	Adding notes with voice interaction
Handling of buttons (D)	2	Information about functions of buttons in voice, Non-reachable buttons controlled with voice
Functions not reachable (D)	2	Control with voice
Artefacts in the truck cab (D)	6	Integrate as much information as possible so that the driver can use voice interaction instead
Remembering activities (P)	7	Voice reminders
Navigation (P)	3	Enable driver to interact with navigation system with voice to get information about break times and spots to stay at
Navigation (S)	8	System that translates foreign road signs and reads the meaning out loud, Voice reminders when close to added locations
Handbook (S)	3	Voice interaction with the trucks handbook
Warnings (S)	4	Get warning difficult to understand explained in voice, Get red warning read out loud
Status checks (S)	4	Be able to ask about road status

3. Help the user find in the truck as a many vehicles look different. One could avoid frustrating scenarios by letting the user say "where is the hand brake?" and thereafter either highlight the hand brake or show where it is located on the trucks screen.
4. Create a program for a fast introduction of the truck. Revenues is what make a company go around, and you get them by selling a product. A so called tutorial system could not only help a novice find in an unknown vehicle, but also promote the truck's different features, making it sell itself.
5. Enable a dialog between the user and the vehicle. If a warning symbol shows on the instrument board, give the user the possibility to ask for more information and possible solutions.

6. Let users communicate with each other. A problem based community could be an effective platform where user's can communicate with each other, giving and asking for advice. By doing this and adding the element of having Volvo experts, one could not only help customers but also get more insight of what people actually want to have in their trucks.

The without question most frustrating problem for 5 out of 7 segments is EU rules. The majority of drivers expressed great frustration towards the fact that they often have to take their required break in the middle of nowhere to avoid having to pay fine for breaking the so called driving and rest act. Many drivers also expressed that their work is very stressful which makes it hard to find the time to properly prepare the day of work. As a matter of fact, some drivers would not even be able to plan their day as they do not get their next destination before they have delivered their current cargo. This resulted in concept two: "I have just started my new job and am going to drive unfamiliar routes. During my first couple of days I have lost a great lot of time taking breaks that have not matched my driving schedule well. Is there any solution for my problem?" The resulting functionality related to this was:

1. Help the user plan his or her route. Warn the driver when he or she has planned a route without the factors below when it will most likely be needed:
  - (a) Rest areas.
  - (b) Places to fill gas.
  - (c) Unavailable roads due to weight and height.
  - (d) Potentially longer roadworks.
  - (e) Roads that tend to form queue.
2. Warn the driver when needed and create a dialog with him or her. Examples of warnings and potential solutions that could be of value are listed below:
  - (a) Your destination is unreachable without refueling. Do you want suggestions on appropriate petrol stations?
  - (b) You will not reach your destination before you have to take your break. Do you want suggestions on appropriate rest areas?
  - (c) There has been a car accident on road X. Do you want to change your route?
  - (d) (The driver has just delivered heavy cargo to a destination) With the vehicles new wight there is a new route to your destination that will save X minutes. Do you want to change your route?

The third concept relates to the truck cabin, and it being a relaxed place where the driver feels at home: "At my previous job I always drove the same truck. Today, however, I switch between three different vehicles and miss the feeling of driving something of my own. Is there any way to get back this feeling with my current situation?" Some suggested solutions were:

1. Attach data to an extern and mobile unit, like a cellphone, as it makes it possible to save information that could be of value to bring into different kind of trucks. Examples of such information is listed below:
  - (a) Adjustable settings within the truck such as chair- and rearview mirrors.
  - (b) Favorite radio stations.

- (c) Saved hot spots such as rest areas, petrol stations, restaurants etc.
2. Make the car feel more personal by giving the user the possibility to customize the car as necessary. Example of such functions are listed below:
    - (a) The car's voice.
    - (b) How warning messages are being presented.
    - (c) The appearance of the vehicles computer, menu's etc.

Some functions did not quite fit under the three concepts above and were not homogeneous enough to form a new one on their own. They are presented below with some general guidelines.

1. Voice control should be activated in two different ways. One way is to push an analog button on the steering wheel as the driver will have his or her hands there while driving (hopefully). The second way should however be activation by speech, like Apple's "Hey Siri!" followed by a command, such as "Hey Siri, what is the time?". This is a crucial feature for drivers who sleep in the car for instance and want to control features such as temperature and light without having to move around in the vehicle.
2. As an addition to the first function in this list, the driver should have the option to use voice control whenever it is needed. As vaguely as this may sound, the main guideline that should be followed is to focus on making features that one could want to control from other places than behind the wheel.
3. Have the option to search for jobs. While this feature requires so much more than a simple interface within the car, one driver would like to be able to search for jobs on the go. For instance, if a driver were to deliver cargo between Stockholm and Gothenburg, he or she could send a message through the car. The function would work like this: "Looking for work. I am arriving at Stockholm in an hour. I have this much space left in my truck and I am going to drive to Gothenburg. Please send a message if you want me to deliver something on the way for you." The car would then transform the speech to text and send it out to a community of people interested in cargo delivery. By doing this one could maximize deliveries and therefore avoid wasting fuel on half empty trucks.
4. Introduce a more advanced status check. As different drivers care more about certain aspects of the car, one should be able to configure a customized status check where each factor will be presented verbally. The driver should also be able to chose what status factors that the truck is going to warn about when for instance running low, and also how it should be presented - either by sound, speech, text or light.
5. As driving the very heavy vehicle that a truck is, adding sensors that will give vocal feedback that the driver could interact with seems to be a safety feature that some drivers want. In addition to a reserve camera, the ability to simply ask the truck if there is something behind or in its blindspots could be of use.
6. When driving in a foreign country, help the driver with translating road signs and give the option to have that specific country's rules and laws read aloud.
7. The last feature presented in this article is something that Volvo cannot fix by themselves, but is something that would have a huge impact on their customer's driving long distances. By using already existing technology in parking lots that

monitor whether a parking space is free or not, driver's would like to be able to ask their car for resting areas where they actually can rest. This would make it possible to say "where is the closest resting area with at least 5 spots available?", and the car would respond with the location and ask the driver if he or she wants to go there.

In addition to the these found features, 70 phone interviews were held with truck drivers with the ambition to find new possible functions (Table 4) while also being able to elaborate the previously found ones (Table 5). Ten test participants were recruited from each of the seven different truck driver segments in order to thoroughly get information from all kinds of truck drivers. Further, this also made it possible to prioritize different functions for different segments as different truck drivers have different needs, see Table 6. The interviews were held by one person who called truck drivers all over Sweden. To make the interviews more structured, an online-based questionnaire was constructed. The interviewer asked the question to the truck driver while simultaneously filling out the questionnaire. Demographical data was also collected during the interviews, see Table 7. These interviews resulted in a better understanding of what speech related functions different drivers want while also finding new and interesting features that may be implemented.

Table 4: New functions

Name	Description
Team based communication	Form a group of trucks that have intercom communication. Be able to share information and see other trucks on map.
Status check from far away	
Conference calls	
Communicate directly with other trucks on the road	
Request help on the road	"anyone close to?"
Socialize with other drivers/find new friends	
Restaurant/rest stop application	"For drivers, by drivers" social media platform
Bicycle detector	Använda tex döda vinkeln för att undvika att krocka med cyklister. "Is there anything coming from behind?"
Activate heat from far away	Using cellphone
Macro based communication	"1 = radio channel 1" etc

Table 5: Elaboration on previously suggested features

Name	Description	Comments
To do list/Notes	"Write down 'Tullgarns slott'"	
Reminders	"Please remind me to water the flowers when I get home."	Should work like iPhones "reminders"
Contact list	Call Filip, where does Christian live etc.	Share contacts, information, locations etc
Media	Control music, radio, podcasts, audio books etc.	-
Warning messages	What does this warning message/symbol mean?	"Give me more information"
Coaching messages	Coaching (Eco driving). "How well am I driving at the moment?"	Statistics is fun!
When do I have to stop?	"You can drive X more hours/minutes before you need to take a break."	-
Get notified when destination is unreachable	When the driver needs to take a break or find a diesel station	-
Activate voice control by using button on steering wheel		
Activate voice control by saying "Hey Volvo!"		Important for functions the drivers' want to control when not driving
Notify when the driver card is not inserted correctly		Not interesting
Order management	Hantera ordrar. Feedback. Koppla till navigering.	Very interesting function
RSS-feed	Chose from news, Twitter accounts etc.	This function was not interesting, since most people get their news by radio.
Status check	Status check	The driver wants to customize what to check
Manual	Digitalize the truck's manual	Very interesting function
Control sunroof		Good when driving / in bed
Control light in coupé		Good when driving / in bed
Control heat in coupé		Can be awkward in practice
What does button X do?	Press a button and ask the truck what it does	Not interesting for experienced drivers
Make some buttons controllable by voice.	Filip knows which buttons would be best to implement!	-
Short introduction of the car	"Please give me information about the navigation system."	Not interesting for experienced drivers
Translate foreign traffic sign		Very interesting function
Ask for new laws and rules when entering a new country		Not interesting for experienced drivers

Table 6: Prioritisation of functions within the different segment. (H) means very valuable, (M) could be useful, (L) not likely to use.

	Inter-regional haul	Demanding long haul	City Distribution	Regional distribution	Light Construction	Heavy Construction	Heavy Transport
Note taking	1 (H)	1 (H)	1 (H)	1 (H)	1 (H)	1 (H)	1 (H)
Handling contacts	2 (H)	2 (H)	2 (H)	2 (H)	2 (H)	2 (H)	2 (H)
Digital manual	3 (H)	3 (H)	3 (H)	3 (H)	3 (H)	3 (H)	3 (H)
Driving improvement (eco-driving)	4 (H)	4 (H)	4 (H)	4 (H)	5 (H)	5 (H)	5 (H)
"You won't be able to reach your destination because of X"	5 (H)	5 (H)	7 (M)	6 (H)	11 (M)	11 (M)	11 (M)
Status check	6 (H)	6 (H)	6 (H)	7 (H)	4 (H)	4 (H)	4 (H)
Order management	7 (H)	7 (H)	5 (H)	5 (H)	9 (M)	9 (M)	9 (M)
"You need to take a break"	8 (M)	8 (M)	8 (M)	8 (M)	10 (M)	10 (M)	10 (M)
Laws and rules	9 (M)	9 (M)	9 (M)	9 (M)	6 (M)	6 (M)	6 (M)
"What does this button do?"	10 (M)	10 (M)	10 (M)	10 (M)	7 (M)	7 (M)	7 (M)
Introduction of functions (tutorial)	11 (M)	11 (M)	11 (M)	11 (M)	8 (M)	8 (M)	8 (M)
RSS-feed (news)	12 (L)	12 (L)	12 (L)	12 (L)	12 (L)	12 (L)	12 (L)

Table 7: Demographics

	City Distribution	Regional distribution	Inter-regional haul	Demanding long haul	Light Construction	Heavy Construction	Heavy Transport
Participants	10	10	10	10	10	10	10
Mean Age	37	42	33	40	46	42	51
Mean Work Experience	18	21	19	24	21	25	24
Gender f/m	0/10	0/10	0/10	1/9	0/10	0/10	0/10
Has slept in the truck	80%	90%	100%	100%	70%	60%	80%
Has driven outside of Sweden	20%	10%	20%	40%	10%	0%	30%
Previous SDS experience	70%	60%	80%	80%	60%	50%	60%
Wants voice in the vehicle	100%	100%	100%	90%	100%	100%	100%

## 2 How should speech interfaces be designed and implemented?

Current speech interfaces implemented in cars on the market today consists of commands for phone, entertainment, navigation, and many also climate control and connected web services (through phone). The commands are initiated by the user but then a system-driven stepwise specification of the request that limits when, how and what information the user can speak takes over. Some systems allows the user to take a short cut by given one-shot-commands if the user can provide all the necessary information in one utterance. The quality of ASR is a big problem and cause for frustration of many users.

Some of these limitations are being addressed and a new improved platform for speech recognition and natural language understanding has been released by Nuance to be used by most car manufacturer in the near future. The introduction of integration of mobile phones with speech functionality, CarPlay by apple and AndroidAuto by Google, also raises the bar of speech functionality in cars. However, even with this development there is still a big gap between commercially available speech interfaces and the full-fledged spoken dialogue systems that has been developed in research systems in the beginning of 21st century. Much can be learned from studies on these systems in how challenges with speech interaction can be addressed, without having to go the whole way with advanced speech and language technology such as deep parsing, dialogue and discourse management, and extensive knowledge representation and reasoning.

### Knowing what, when and how to speak

A big problem with services provided by speech interfaces is that users do not know the commands or requests that can be made, i.e. what functionality the system provides. Or they do not know the correct words or phrasing to use or what is the correct format for data entry. This leads to frustration when the user cannot get the information that is available. There are several strategies that can be used to improve on this.

A general guideline for speech systems is that the system should match the input vocabulary and grammatic complexity to the output the system gives, as to set the right user expectations. When speech is combined with a graphical interface the words shown on the screen should also be possible to use for spoken interaction. Another way of setting the right user expectations is to match the quality of the system voice to the quality of the speech recognition, a more "robotic" voice can lower user expectations while a very human-like voice will raise them.

When designing the system it is important to gather information on how users expect to use the system, for example through Wizards of Oz studies. The structure of the services provided should be simple and match the users mental models. It is, for example, important to consider what commands should be global and what are context dependent and how to convey the difference between these to the user. The vocabulary used for commands and request must also match the language used by users. If commands or data entry differ from what users would expect it is important that the system clearly express this and motivate why it is so. To help the user the system should provide informative prompts stating both what information is requested and the format expected. If information can be given in different formats the system should indicate this.

To help the user learn what can be said and how it should be expressed the system can coach novice users and provide more frequent and more detailed prompts. When



the user has become an expert the system should adapt the interaction to allow for more efficient interaction. A solution can also be to allow the user to set up a profile stating both what type of information that he will need and what type of interaction he prefers.

## **Error recovery**

Since automatic speech recognition is not perfect, and the car is a problematic environment for speech recognition, the system will need strategies to handle problems with imperfect recognition and potential mis-recognition. Usually the system has a threshold for when an input is deemed understood and if the ASR do not reach this threshold the system states that the user input was not understood and ask the user to repeat himself, sometimes with the added request for the user to rephrase the utterance. A problem is that the user do not know it is the ASR that failed or if the user provided the wrong command or wording. For inexperienced users this can lead to correct commands being perceived as faulty and not used by the user again. When the threshold is reached the system can still potentially have misunderstood the user and therefore some systems choose to always repeat the understood command or input as a confirmation strategy. This can however lead to very stilted and tedious interaction.

A more sophisticated strategy is to use thresholds to classify utterances as accept, confirm and reject and only confirm some utterances. The confirmation can be done in an implicit way as part of the prompt, which gives the user the possibility to correct if necessary but otherwise continue the interaction. The thresholds can also dynamically change for different contexts depending on the semantic content of the recognised words. The current context can even be used to set a specific recognition grammar for the ASR.

## **Dialogue handling**

The dialogue model most commonly used in cars today can lead to long and tedious dialogues where the user must wait through system prompts, press the push-to-speak button and then wait for a signal from the system that it is ready to hear the users input, before being able to speak. There are several things that can be done to increase the naturalness of the interaction and a more human-like dialogue. One is that the system is always listening and that the user do not have to press a button and wait for a system ready indication. Another is that the turn-taking between system and user should allow for barge-in where the user can skip a system prompt, preferably by speaking over it but otherwise by pressing the push to speak button.

The system-driven interaction should be replaced with real mixed-initiative dialogue where the user can take the initiative and steer the interaction by changing the topic or ask for clarifying information if needed. The user should also be able to overanswer questions and give more information than the system has asked for. A way of doing this is to use an Information State, i.e. a feature value structure, that keep tracks of the current request/command and what information pieces are needed. This allows for a much freer interaction where information can be provided independent of order, in contrast to state chart based approaches. It also support the system when the user provides ambiguous or vague information and the system need to pose clarification questions. The system can asks questions based on missing slots until a complete and unambiguous request has been specified. For example, for destination entry a city and street address are mandatory but additional information like state/province, zipcode etc can be used for disambiguation.

## **Multi-modality**

Many of today's systems combine speech and touch screen thus allowing for multi-modal input and output. Since screens typically are more distracting than speech they should be used to provide redundant information, allowing the user to use speech only. Redundancy may also be important if speech cannot be used in some situations and then the manual interface should provide the same services as the speech system. In best cases the user can combine modalities freely and use speech together with manual input, for example pointing in a map and saying "I want to go here".

Speech versus visual-manual interfaces in  
trucks: effects on driver distraction, user  
acceptance, and perceived efficiency

Sofia Lindvall

22 May 2016



## Abstract

Truck drivers often have a tight time schedule and therefore need to carry out several in-vehicle tasks while driving, such as making phone calls, writing down information and navigating to new places. Performing these tasks using a visual-manual interface can impose visual distraction which has been shown to lead to safety-critical events on the roads. Instead of using a visual-manual interface, a speech interface could be a safer alternative if designed properly. However, the cognitive load demanded by speech interfaces and the connection between cognitive load and driving behaviour is not fully known. In this study, a speech interface and its visual-manual counterpart were evaluated and compared in terms of visual distraction, cognitive load and user efficiency and perceived acceptance. Eye tracking was used to measure visual distraction. The measurements used for cognitive load were the Tactile Detection Response task (TDRT) and the Driving Activity Load Index (DALI). Perceived acceptance and efficiency were measured using the System Usability Scale (SUS), the Subjective Assessment of Speech Systems (SASSI) and semi-structured interviews.

The conclusions were that (1) the speech interface was less visually distracting than the visual-manual counterpart, (2) the speech interface was less cognitively demanding than the visual-manual interface, especially in the navigation task, (3) the speech interface was safer to use while driving compared to the visual-manual interface and (4) the speech interface had higher user acceptance and efficiency than the visual-manual interface. Further research should investigate the connection between cognitive load and driving behaviour, such as lane keeping and brake response time, by employing a variety of speech tasks with various complexity as well as including speech interfaces entirely free from visual demand. The focus should be on the differences between baseline driving and speech interaction, as opposed to speech interaction and visual-manual interaction.

## Acknowledgements

First, I would like to thank Annika Silvervarg and Arne Jönsson from Linköping University, as well as Ida Esberg and Christian Jernberg from Volvo Trucks for sharing their knowledge and giving valuable feedback throughout this work. I have really enjoyed working in this project.

I would also like to thank the truck drivers that participated in the study for putting up with the tests and sharing their opinions.

Finally, I would like to thank Sam Thellman for encouragement and providing the best support anyone could wish for.

# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	Aim and research questions . . . . .	2
1.2	Limitations and delimitations . . . . .	3
<b>2</b>	<b>Background</b>	<b>4</b>
2.1	Cognitive load and attention . . . . .	4
2.2	Driver distractions and accidents . . . . .	5
2.3	Speech interaction in vehicles . . . . .	8
<b>3</b>	<b>Method</b>	<b>10</b>
3.1	Study design . . . . .	10
3.2	Measuring cognitive load . . . . .	11
3.3	Measuring visual distraction . . . . .	13
3.4	Measuring user acceptance and perceived efficiency . . . . .	14
3.5	Pilot study . . . . .	16
3.6	Participants . . . . .	17
3.7	Procedure . . . . .	17
3.8	Data analysis . . . . .	18
<b>4</b>	<b>Results</b>	<b>20</b>
4.1	Cognitive load . . . . .	20
4.2	Visual distraction . . . . .	27
4.3	User acceptance and efficiency . . . . .	36
4.4	Summary . . . . .	40
<b>5</b>	<b>Discussion</b>	<b>42</b>
5.1	Results . . . . .	42
5.2	Method . . . . .	44
<b>6</b>	<b>Conclusions and further research</b>	<b>48</b>

# 1 Introduction

The trucking industry plays a huge role in our society, delivering goods for consumers and manufacturers throughout the world. There are many different types of truck drivers and the needs and goals of the drivers will differ. But what they all have in common is that they need to deliver in time, which often leads to a stressful working environment with a tight time schedule. Because of the time limits they often need to carry out secondary tasks while they are out on the road driving, as they do not have time to stop the vehicle every time they need to make a phone call or navigate to a new address. It is important to note that the tasks they need to carry out will differ depending on what type of truck driver they are. For example, long haul drivers who spend weeks in the truck while driving long distances to new locations, or city distributors who drive the same routes every day. This leads to different goals and needs for the systems used for secondary tasks. Some examples of secondary tasks are to communicate with customers and colleagues, get navigation assistance, to retrieve truck information, and note down reminders. Carrying out these tasks using an interface that requires looking at the interface or taking the hands away from the steering wheel, which is typically the case for visual-manual interaction, could direct the driver's attention away from the driving task. Distractions imposed on drivers have been shown to affect driving behaviour negatively as well as being a factor leading to accidents (NHTSA, 2012; Wilson and Stimpson, 2010). The type of distraction with the most obvious connection to negative affects on safety is visual distraction (Hickman et al, 2010; Engström, 2011; Klauer et al, 2006), which organisations such as the National Highway Traffic Administration (NHTSA) and The International Organization for Standardization (ISO) recommend to avoid. By introducing a well-designed speech interface the visual distraction when carrying out secondary tasks while driving can be reduced. But even though the visual distraction is less with a speech interface the cognitive distraction imposed by using the interface can still be high (Lo & Green, 2013). This is something that need to be taken into consideration when introducing a new interface in the trucks.

The International Organization for Standardization (ISO) is currently working with developing a standard for the use of a Detection Response Task (DRT) to measure the effects of cognitive load on attention for secondary tasks involving interaction with visual-manual, voice based or haptic interfaces (ISO 17488, 2015). In addition to that, the National Highway Traf-



fic Safety Administration (NHTSA) is also working with minimising cognitive load imposed on drivers from, for example, in-vehicle and portable devices (Ranney et al, 2014). NHTSA is therefore working on guidelines that can be followed in order to avoid distraction for drivers.

However, the connection between cognitive load and real crash risk is not conclusive. Some researchers have found that cognitive distraction is indeed decreasing safety on the road. For instance, a recent study found voice systems to cause cognitive load that lasted even after the voice task was completed, which could mean that the drivers need time to establish situational awareness (Strayer, 2015). However, some research results show that a higher cognitive demand for the drivers results in a safer driver behaviour (Victor et al., 2014; Baron and Green, 2006).

Because of the possible benefits of interacting with voice instead of an interface requiring visual-manual interaction, we implemented and evaluated a prototype of a speech interface in a truck. The functionality was based on user research conducted in an initial stage of the research. The speech interface and a visual-manual counterpart was evaluated in terms of visual and cognitive distraction as well as user acceptance and perceived efficiency.

## 1.1 Aim and research questions

The aim of the thesis is to evaluate and compare a speech interface to its visual-manual counterpart in terms of distraction, safety and the drivers acceptance and perceived efficiency of the interfaces.

The research questions investigated are:

- How does the speech interfaces as compared to the visual manual interface affect distraction while driving?
- Is interaction via voice a safer way to communicate compared to the visual-manual interface?
- What is the truck drivers acceptance and perceived efficiency of the speech interface compared to it's visual-manual counterpart?

## **1.2 Limitations and delimitations**

The aim of the thesis is not to make conclusions of speech interfaces in general, as those could differ greatly depending on the complexity of the tasks and the design of the interface. The data collected is based on truck drivers employed as test drivers and should not be seen as a representation of truck drivers in general. This thesis has focused on visual and cognitive distraction and its connection to safety. Other measures would also be relevant to use, but could not be included in this study.

## 2 Background

This section contains a description of theories about cognitive load, attention and driver distraction as well as its connection to accidents. It will also discuss how speech interaction can be used as a safer alternative compared to visual-manual interaction in vehicles.

### 2.1 Cognitive load and attention

In order to understand cognitive load and its impact on interaction with in-vehicle systems we need to understand how the working memory is organised. The working memory holds the activated portion of the long-time memory and moves that in and out of the short-time memory. Alan Baddeley (2006) suggest that the working memory consists of four elements: the visuospatial sketchpad, the phonological loop, the central executive and the episodic buffer. The visuospatial sketchpad holds images, the phonological loop holds inner speech for verbal comprehension and acoustic rehearsal, the central executive coordinates by deciding what information to process and how to process it. The episodic memory integrates memories from the different systems to an episodic representation. One of the major assumptions of the cognitive load theory is that the working memory only has a limited amount of resources (Bannert, 2002; Ayres and Paas, 2012, Young and Stanton, 2002). When the demand exceeds the available resources, the performance of a task will degrade.

There exist several theories about attention and how this is managed when we are presented with several tasks, which is called divided attention (Sternberg, 2009). Theories have moved towards a common view of limited attentional resources, which refers to a fixed amount of attention that can be allocated according to what the situation demands (Sternberg, 2009). However, more recent theories claim that this is an oversimplification and that dividing attention is easier when the attention is distributed over different modalities, the so called *multiple resource theory* (Wickens, 2002: ISO 17488, 2015).

Another theory that exists is the Malleable Resource Theory (MART) developed by Young and Stanton (2002). While the previously presented theories of attention claims that we have a limited amount of resources available, MART claims that the size of available resources can change depending on the task. This could mean that reducing demand does not have to lead

to an improvement in performance of a task. MART proposes that resources may shrink to accommodate the demand required by the task and that this could lead to a degradation of attention and performance in tasks that don't require much demand. The consequences of reducing demand of a task could lead to a driver having difficulties handling a safety-critical event. Therefore, a secondary task requiring cognitive load might increase safe driving behaviour by increasing available resources. The theoretical ground used will affect conclusions on how secondary tasks should be designed to avoid driver distractions and accidents.

## 2.2 Driver distractions and accidents

Truck drivers as well as all other drivers are exposed to several distractions when out on the road driving. Some of the distractions can in some cases cause safety-critical incidents or even accidents. It is therefore important to be aware of the different types of driver distractions that exists and how to avoid them when designing in-vehicle systems.

The AAA Foundation for Traffic Safety (AAAFTS) defines driver distraction (Stutts et al, 2001: 6) by claiming it to occur when:

*'a driver is delayed in the recognition of information needed to safely accomplish the driving task because some event, activity, object, or person within or outside the vehicle compels or induces the driver's shifting attention away from the driving task'.*

Lee et al. (2008: 38) focus on limited resources when defining driver distraction:

*'diversion of attention away from activities critical for safe driving towards a competing activity'.*

The National Highway Traffic Safety Administration (NHTSA) (2012: 5) defines driver distraction as the following:

*'a specific type of inattention that occurs when drivers divert their attention away from the driving task to focus on another activity'.*

All definitions highlights that driver distraction occurs when the drivers are shifting attention away from the driving task. The AAAFTS and Lee et al. (2008) include the concept of *safety* in their definitions meaning that distraction derives when the inattention is critical for safe driving or effects recognition of information needed to being able to drive safely.

Several studies have shown that inattention plays a key role in vehicle crashes (Klauer et al., 2006; McEvoy et al., 2005; Stutts et al.; 2001, Engström, 2011). Inattention can derive from several causes of distraction, such as engagement in a secondary task, fatigue, non-specific eye glance and driving related inattention to the forward roadway (Klauer et al., 2006). NHTSA (2012) describes three types of driver distractions which are visual, manual and cognitive distractions. Visual distraction occurs when the drivers need to look away from the roadway to obtain visual information, manual distractions comes from task were the drivers need to use a hand and therefore remove it from the steering wheel and cognitive distraction occurs when the drivers need to take mental attention away from the driving task. NHTSA (2012) recommends that distracting tasks interfering with the driver's ability to operate the vehicle in a safe way should be avoided. The following are some examples of tasks that should not be carried out while driving (NHTSA, 2012: 9):

*'Displaying images or video not related to driving; displaying automatically scrolling text; requiring manual text entry of more than six button or key presses during a single task; or requiring reading more than 30 characters of text (not counting punctuation marks).'*

According to NHTSA (2012) 17% off all crashes reported to the police involve driver distraction. Furthermore, in the US year 2010, for 3% of all crashes it was explicitly stated that the driver was distracted using an integrated device and 5% a cell phone. Wilson and Stimpson (2010) investigated trends in distracted driving resulting in fatalities. They analysed data from the Fatality Analysis Reporting System (FARS) that records all data regarding road fatalities in the US from 1999 to 2008. The use of a regression analysis predicted that the increasing volume of texting resulted in 16 000 additional road fatalities from 2001 to 2007. Deaths caused by distraction increased from 10.9% in 1999 to 15.8% in 2008, with much of the increase after the year 2005. Deaths related to distracted driving increased 28.4% from

2005 to 2008. Pickrell (2015) studied mobile phone use while driving using the National Occupant Protection Use Survey (NOPUS). Pickrell found that 4.6% of the drivers talked in their phone by holding it against their ear, 0.5% spoke with a visible headset and 1.7% of the drivers manipulated handheld devices while driving. Young and Lenné (2010) analysed results from a self-reported internet survey of 287 drivers in Victoria, Australia regarding their engagement in distracting activities. They found that almost 60% of the drivers use a mobile phone when driving and over one third used the phone handheld. They did also find that drivers try to modify their behaviour when conducting a distracting activity, such as reducing speed, increase distance to the vehicle in front and stopping the vehicle. However, this might not be an option for truck drivers with a tight time schedule.

The connection between different distractions and safety needs to be further explained. Hickman et al. (2010) collected naturalistic data from commercial trucks and buses to study what in-vehicle tasks are performed that are connected with real-world traffic danger. The results were that talking and listening in a mobile phone was not associated with an increase of the odds of being in a 'safety critical event'. However, they could observe a very strong relationship with being in a safety critical event and texting, accessing the internet or e-mailing while driving. Thus, the risk of being involved in a safety critical event was connected with visual distraction. Those tasks that had the highest visual distraction was also the tasks which had the highest risk. The risk of being involved in a safety critical event and the connection to visual distraction has also been found by others, for example Engström (2011) who found a connection between being in a safety critical event and the time spent with eyes off the road. Furthermore, Victor et al. (2014) found that the crash risk was high for texting, which is a visually demanding task. They mean that an implication of these results is to design interfaces that are as less visually distracting as possible. They also emphasize the potential for interfaces that are non-visual. This could mean that voice interaction is a safer alternative if designed to avoid being visually demanding.

In regards to risks associated with tasks requiring cognitive load, different studies have shown different results. In some studies, memory tasks actually have shown to decrease risks; talking in a hands-free phone were found to significantly reduce the risk for a safety-critical event to occur compared to not having a phone conversation (Engström, 2011; Victor et al., 2014) and lane variability have been found to decrease (Engström, 2011). The section below will further discuss cognitive load and driver behaviour in relation to

*speech interfaces.*

### **2.3 Speech interaction in vehicles**

The use of speech interaction instead of the visual-manual counterpart is a way of reducing visual distraction for the drivers while out on the road driving. But the demand the use of speech interfaces has on cognitive load and how that correlates to risks in traffic is unclear.

Strayer, Cooper, Turrill, Coleman and Hopman (2015) examined implications on cognitive load when using speech interaction while driving. They selected tasks that were free from visual demand and measured cognitive load with the Detection Response Task (DRT), the NASA TLX survey and video recordings. Their study found a significantly higher cognitive load when the driver was out on the road driving and at the same time carrying out the voice tasks compared to when just driving. Strayer et al. (2015) did not perform a comparison with a visual-manual counterpart, but added the highly cognitive demanding operation span (OSPAN) task in an auditory variant (the OSPAN task is based on letting the participant solve mathematical problems and at the same time remember words which should later on be recalled). The results of cognitive workload between the OSPAN task and the voice tasks did not differ, meaning that the voice task imposed a high cognitive demand on the drivers. Another finding from the study was that the DRT data recorded exhibited that the cognitive load caused by interacting with the voice system lasted up to 18 seconds after the interaction. Strayer et al. (2015) explained this by proposing that the drivers need time to establish situational awareness.

However, several results points in a direction of voice interaction being a safer alternative in the automotive domain. Baron and Green (2006) found several advantages with using a speech interface while driving. For example, the use of a speech interface resulted in fewer lane departures, steadier speed, less workload (based on a subjective measurement) and less glances away from the roadway. According to Lo and Green (2013) the level of distraction is lower when using a speech interface and the speed for task completion is quicker in some tasks such as entering an address while navigating. They also claim that speech interfaces makes the driver keep lanes better, shortens the reaction time of events happening in the periphery and results in the driver glancing away from the roadway fewer times. Even though the subjective workload has been shown to be less in some studies, speech interfaces with

poor speech recognition accuracy could impose a high cognitive demand on the drivers (Lo & Green, 2013). Lee et al. (2001) conducted a car following task to see how speech based e-mail systems affects drivers response to braking of a lead vehicle. They compared a baseline with two different e-mail systems: one simple and one complex. They found a 30% increase in reaction time when using a speech system compared to when just driving. Subjective measurement of cognitive workload using NASA-TLX showed that the use of an e-mail system had significantly higher rating for workload compared to when only driving and that the complex system was significantly higher rated than the simple version. This shows that there can be differences between different types of speech based systems, depending on their complexity.

Another aspect to consider is that speech interfaces not necessary are free from visual-manual demands. Reimer and Mehler (2013) conducted an on-road study where they could confirm this. Their findings show that implementations of voice interfaces that are multi-modal can result in visual-manual demands where the interface, for example, require the user to view a display several times to be able to add information. Speech interfaces that are not well-design might impose both visual and cognitive demand on the drivers. But one question is whether it substitutes as a better alternative compared to visual-manual interfaces. He et al. (2014) compared texting while driving and interacting with speech-based text entry versus a handheld-cell phone in a car following task. They found that both interactions decreased the driving ability compared to the drive-only condition in regards to, for example, more speed variation, increased brake response time and increased variation in gap distance. However, the speech-based interaction was not as bad for the driving performance as the handheld-device was. Their conclusion was that speech interaction might be better, but still not entirely free from hazard.

Based on this background, speech interfaces need to be evaluated both in terms of visual-manual distraction as well as cognitive distraction, in comparison to just driving and using visual-manual interfaces.



## 3 Method

This section describes the method chosen and the procedure of the study and data analysis.

### 3.1 Study design

#### The interfaces

The speech system is implemented for Swedish and English. When the user wants to interact with the system, he or she presses a push-to-talk button placed on the right arm rest. The system signals that it is listening with a sound and visual information on the cluster display. The system gives both auditory and visual feedback to the driver, but all tasks are possible to carry out without viewing any displays. For the visual-manual system, a secondary display is located to the right of the steering wheel where the driver could assess his or hers mobile phone, the navigation system and the entertainment system.

#### Set-up

The study design was a repeated-measures within-subjects design, meaning that all participants performed all combinations of tasks and interfaces. A baseline was included with data from when the participants were just driving, without interaction with the interfaces. The study was chosen to be carried out on Hällered's proving ground on a 6.2 kilometre long oval motorway. The weather conditions were dry, but two of the participants had to drive in strong sunshine. They were asked to drive at a speed of approximately 80km/h. No help systems were allowed to use while driving.

#### Secondary tasks

For both the speech interface and the visual-manual counterpart, the following tasks were carried out:

1. Call your own phone number. Then call X from the phone book.
2. Play Madonna, Like a prayer. Then ask the system to remind you to post the Declaration of income to the Tax Agency.

3. Navigate to Vasagatan 15, Stockholm.
4. Tell us the next time you need to take a break.
5. Check your warning messages, vehicle message 2.

For task 2, the visual-manual counterpart was to write the reminder down on a piece of paper. In order to get sufficient data for analysis, short tasks were made longer by putting them together such as those in task 1 and 2, in order to get at least 5 DRT data points, which is needed for the analysis (ISO 17488, 2015).

## 3.2 Measuring cognitive load

This section describes the objective and the subjective measurement of cognitive load that have been used in the study.

### The Tactile Detection Response task

The Detection Response Task (DRT) task is carried out by repeatedly presenting simple targets and recording the driver's response time (ISO 17488, 2015). There are different types of DRTs available which differs in regard to what stimulus they present, visual, auditory or tactile. With Tactile DRT (TDRT), which is the selected variant in the study, the driver has an electrical vibrator taped on the left shoulder. According to ISO 17488 (2015) an advantage with TDRT is that the target remains in the same position relative to the driver's eye position, which eliminates the variability between the target and the head position. Another advantage is that TDRT does not give a visual stimulus. This eliminates conflict between detection of a visual target and the visual demand of driving, which according to Engström (2010) could mean that the TDRT is the purest variant for measuring attentional demand. ISO 17488 (2015) specifies that TDRT could be preferable if a voice-controlled interface requires glances away from the roadway. The reason for this is that the TDRT bypasses the visual modality and therefore has the highest specificity for attentional effects on cognitive load.

Before performing the TDRT data collection training should be performed for all participants in order to reach a steady performance during the test. The recommended order described in ISO 17488 (2015) is the following: (1) the secondary tasks under evaluation, (2) the DRT, (3) the primary (driving

or driving-like) task (if used in the study), (4) the tasks together. The tasks used in the practise trial should have the same complexity as the one used in the study, but they should not be the same. An example could be that the practice trial and the study have the same length of a street name when entering an address in a navigation task, but it should not be the exact same address. During both the training and tests, the experimenter should give the participant assistance and coaching if difficulties with the task occurs, so that the task is completed in an appropriate manner. DRT training keeps going until the experimenter thinks that the participants responds to the stimulus in a stable manner and the participant feels comfortable. The experimenter should observe the participant to assure that he or she is trying to respond as quickly as possible, as well as assuring that the participant is not just clicking the button without perceiving a stimulus. If these behaviours occur and the participant is not able to change that, they must be eliminated from the study. (ISO 17488, 2015).

Recommendations from ISO 17488 (2015) were followed to specify the set-up for stimuli and response for the DRT. The max stimuli duration was 1 sec and the stimulus was turned off at the same moment the participant responded. The stimulus cycle period refers to the time from the beginning of one stimuli to the beginning of the next stimuli. This varied between a uniform distribution (a known number of outcomes equally likely to happen) of random values from 3 to 5 seconds. The participants got a micro-switch attached to the index finger, the middle finger or the thumb on the left hand to click on when they perceived the stimulus.

### **Driving Activity Load Index**

The Driving activity load index (DALI) is a questionnaire for a subjective evaluation of mental workload especially developed for the driving context (Chin et al, 2004). DALI is based on NASA-TXL in which mental demand, physical demand, temporal demand, performance, frustration level and effort are factors that are taken into consideration. According to Chin et al. (2004), DALI was created by developing NASA-TLX to better fit drivers in a vehicle equipped with an in-vehicle system. The questionnaire (see Appendix I) is taking into consideration and evaluate task demands, effort of attention, interference and stress. Task demand includes visual, auditory, tactile and temporal demands. The results from DALI is normally used to compare the results from a normal driving situation with a situation where

the driver's workload is influenced. In this study, DALI has been used to measure subjective mental workload for three conditions: a baseline, when interacting with the speech interface and when interacting with the visual-manual counterpart.

### 3.3 Measuring visual distraction

A common way of measuring the driver's visual distraction is to use eye tracking technology, which is used to measure the position of the driver's eyes relative to the road or other areas of interest (McGehee, 2014). Eye tracking has been used in several research areas such as human factors and human-computer interaction (Bergstrom and Shall, 2014). Looking at areas of interest can be used to analyse different components of a task and can serve as an objective complement to other measurements. Fixations are a common measurement for eye tracking data, which shows where the participant is fixating his or her eye gaze. However, a question is whether the eye gaze reflects where the person is locating his or her attention. The eye-mind hypothesis means that where the person looks indicates where the person's attention is allocated (Ghaoui, 2005). Just and Carpenter (1976) proposes that fixations reflect what is at 'the top of the stack' and that the processing time of the task is connected to the fixation time. Also Yarbus (1967: 190) means that seeing is linked to cognitive goals:

*'Eye movements reflect the human thought - processes so the observer's thought may be followed to some extent from records of eye movements'*

However, this has been questioned. Greene, Liu and Wolfe (2012) conducted a study where they let the participants look at a picture showing a family and were asked to carry out different tasks such as give the ages of the people or remember the clothes they wore. Greene, Liu and Wolfe (2012) were not able to predict what task the participant conducted by looking at their eye gaze. When measuring eye gaze in driving studies, it is important to be aware of the different theories available.

In this study, amount of fixations, number of glances and number of glances greater than 2.0 seconds were analysed to compare the glance behaviour for three different conditions (baseline, speech and visual-manual interaction). The reason for analysing glances with a duration of 2 seconds or more is that NHTSA (2012) has found that they are connected with an

increase of crash risk and therefore should be avoided. The areas of interest used were the road and the task displays located in the truck. The equipment used was Ergoneers.

### **3.4 Measuring user acceptance and perceived efficiency**

This section describes the System Usability Scale, the Subjective Assessment of Speech Interfaces questionnaire, and the interviews.

#### **System Usability Scale**

The System Usability Scale (SUS) is a tool developed for quick and easy evaluating of a systems overall subjective assessment of usability (Brooke, 1996). According to Brooke (1996) an overall assessment of a systems usability and how it compares to other systems is often what is needed when evaluating a system. Using objective measurement of usability can be difficult, for example, task completion time can differ greatly from system to system without affecting the usability. However, a subjective measurement can be compared between systems, even systems used in different domains. According to Bangor, Kortum and Miller (2008), SUS can be used for different products and services, from voice based interfaces, to web sites and hardware platforms.

The SUS is a Likert scale consisting of 10 items normally ranked from 0-5 (Brooke, 1996). In this study a ranking of 0-7 was used. According to Brooke (1996) the scale is generally used after the respondent has interacted with the system and before any discussion about the system has been carried out. After the data collection a SUS score is calculated. For half of the items (those with odd numbers) the score is the scale position minus one. For the other half that are phrased negatively (the even numbers) the scale position is subtracted from 5. The sum of all scores is then multiplied by 2.5 (or 1.67 if used on a 7 point scale) to get a score between 0-100. However, the use of both positive and negative items in questionnaires has been questioned. Sauro and Lewis (2011) found no evidence of biases when using only positive items. Based on this, all positive items were used in this study (see Appendix II for all questions).

Bangor, Kortum and Miller (2008) collected data from 200 studies using SUS and found that the mean SUS-score for all studies was 69.69 (SD = 11.87). They mean that products with at least passable scores have SUS

scores above 70 and that superior products score better than 90. If the product have a score below 70 it is recommended to be improved.

### **Subjective Assessment of Speech Interfaces**

Hone and Graham (2000) developed a tool in the form of a questionnaire for a subjective evaluation of speech system interfaces. The tool is called Subjective Assessment of Speech Interfaces (SASSI). It has 50 items on which the participants rates their agreement on a Likert scale (7-points in this study). The items are balances so that every other is positive and negative.

The items are divided into six main factors: System Response Accuracy, Likeability, Cognitive Demand, Annoyance, Habitability and Speed. System response accuracy measures if the users perceive the system as accurate and if it is doing what's expected. According to analyses made by Hone and Graham (2000), this could be a particular important aspect of interaction with a speech recognition system. Likeability refers to if the users thinks the system is useful, pleasant and friendly. Cognitive demand measures the perceived amount of effort that is needed to interact with the system. Annoyance measures for example how boring, irritating and frustrating the system is perceived as. Habitability refers to if the users know what to do as well as if the system knows what it is doing. Speed is how fast the system responds to the users input. (Hone and Graham, 2000)

In this study, SASSI was used as a complement to SUS. In order to avoid repetition and the time required to fill in the questionnaires, the number of SASSI questions were reduced (see Appendix III). SASSI questions overlapping with SUS questions were removed, and questions from the SASSI factors Annoyance and Speed were included as they had no corresponding question in SUS.

### **Interviews**

Conducting interviews in the setting of which the product will be used is preferred, as this will provide memory cues for the participants while answering the questions (Goodwin, 2009). Because of this the interviews were carried out in the truck directly after the driving session. It was a semi-structured interview with five main questions. The number of questions were cut down because of time constraints. The purpose of the interviews was to complement data from the questionnaires to get insights about the truck driver's

goals, needs and mental models. The interview was conducted with 11 of the participants and were recorded. The questions were the following:

1. Did you feel understood by the system?
2. Did you always know what to say to the system?
3. What is best with the system?
4. What is worse with the system?
5. What do you wish you could do with the system?

### **3.5 Pilot study**

A pilot study was conducted to find eventual improvements of the study design. The participants chosen were employees at Volvo. All had received driver's licences in order to work with development of the trucks and were novel drivers. An on-road set up was chosen, which means that the participants drove on an open road. The vehicle that was used was a heavy truck which the participants drove in an urban area and on a motorway. The road conditions were mostly dry, but some of the participants drove parts of the experiment in rainy weather. The traffic density differed but was mostly moderate. It was daylight with clear visibility for all participants.

Before the study, the participants received a letter with information about the purpose, the expected duration of the test, and a clarification that safety is the main priority. The participants were instructed to mainly prioritise the driving task, but also the secondary tasks and the DRT. An instruction to respond as soon as the DRT stimulus was perceived was also given.

In line with recommendations from ISO 17488 (2015) all participants received training before the study started. Because of time constraints, the training procedure in the pilot study was modified from the one recommended in ISO. The participant first practiced the speech tasks, the visual-manual tasks and the DRT while standing still. They then drove a test drive to feel comfortable behind the wheel. Next, they started driving with the DRT active for a couple of minutes, which was the baseline condition. The participant was asked to stop at a gas station to fill in the DALI questionnaire. When they drove off again they practiced the speech tasks while driving and then performed the speech tasks. Next stop was to fill in the DALI, SUS and

SASSI questionnaires and after that practice and carry out the visual-manual tasks. The participants were asked to drive back to the starting point to fill in the questionnaires again.

A lesson learned from the pilot study was to make the tasks longer by adding several subtasks into one task. A reason for making the tasks longer is to get more data points from the DRT in order to perform the analysis, as it needs at least 5 data points. The short tasks in the pilot study did not reach up to that amount. It was also found that the training should be more extensive in further studies. When compromising the training it was not enough for the participants to reach a similar level of knowledge for both interfaces. Time points should be noted for when the tasks starts and ends, so that irrelevant data (such as DRT from when the participant is receiving instructions) can be sorted out for the analysis. A detailed guide including all steps and instructions must be made to ensure that all participants receive the same instructions.

### **3.6 Participants**

There were 14 participants in the study, all of which were men. Their age ranged from 27 to 59, with a mean age of 46.6 (SD=10.27). The participants were employed as test drivers at Hällered's proving ground with C/CE driving licenses. Almost all of the drivers used smartphones several times during a day and a navigation system a couple times a month. Two of them had previous knowledge of using a voice system, using it once or twice a week and once or twice a month respectively. All had Swedish as their mother tongue. None had any earlier experience of the DRT.

### **3.7 Procedure**

Upon arrival, participants received a letter with information about the purpose, the expected duration, their rights as participants, and a reminder to always keep safety as the main priority. The participants were then asked to fill out a form regarding background information such as age, gender, and what driver's licences they hold.

Before the driving session, each participant carried out a training session in the vehicle while standing still. Half of the participants were instructed to start with the visual-manual tasks and the other half to start with the speech tasks. The participants received information about the tasks from



the test leader and then got to practice until they performed consistently, with assistance when needed. The training tasks had the same complexity as the test tasks, but with different content. Next, the participant was asked to drive to the test track.

When arriving at the test track, the DRT and the eye tracking started. The participant began with driving a baseline distance for three minutes without carrying out any of the tasks. After the baseline drive, they were asked to stop and fill out the DALI questionnaire when they felt comfortable doing so. They drove off again to start training the voice tasks and the visual-manual tasks while driving. The participant was then informed that the test started. They were instructed to carry out each of the five tasks using one of the interfaces (visual-manual or speech). Then, they were asked to stop to fill out the questionnaires (DALI, SASSI and SUS). Next, they were asked to drive and carry out the five tasks with the other interface. At last, they were asked to stop to fill out the questionnaires again. The participants were then asked to drive back to the starting point where a short interview was conducted.

### 3.8 Data analysis

The first step in treating the data was to remove invalid responses for the DRT data. Based on guidelines, all responses outside the interval of 100-2 500 ms from the stimulus were removed as well as responses that were repeated within the interval (ISO 17488, 2015). This was made to handle possible coping strategies from the participants. Two performance measures were used: response time and hit rate. Response time is the time it takes for the participant to respond to the stimuli after onset. Hit rate is the amount of correctly detected stimuli of all stimuli that were presented.

The first step in the analysis was to investigate if there were any outliers in the DRT data. When viewing a boxplot in SPSS two outliers were identified. When looking closer at the circumstances for the outliers it was noted that one participant was exposed to strong sunshine during the task, which made it very hard seeing the roadway. The outlier was therefore changed to the second highest value plus one, according to a method from Fields (2009: 153). The same procedure was carried out for the other outlier. After the transformation of outliers the assumption of normality was satisfied, as assessed by Shapiro-Wilks test,  $p > .05$ .

As for the hit rate data, a boxplot revealed three outliers. A closer ex-

amination of the outliers could not reveal a reason such as issues with the equipment, weather conditions or that the data from the participant was deviant in several conditions. By assessing Shapiro-Wilks test for the hit rate data it was revealed that data was not normally distributed,  $p < .05$ . According to ISO 17488 (2015) hit rates generally have a high ceiling effect for especially baseline data and in those conditions non-parametric tests should be used instead (ISO 17488, 2015: 12-13). The outliers were kept in the analysis with the argument that non-parametric tests are more robust dealing with outliers.

The eye tracking data had outliers, as assessed by inspection of a boxplot for values greater than 1.5 box-lengths from the edge of the box. One extreme outlier (more than 3 box-length away from the length of the boxplot) for the baseline condition was due to a calculation error, and was changed to the next highest value + 1. The other outliers were kept, as no reason such as equipment errors, weather conditions or a participant with overall deviant data could be identified in the data collection. When testing the data for normality using Shapiro-Vilks test as well as viewing histograms, it was revealed that the data was not normally distributed,  $p > .05$ . The choice was therefore to use non-parametrical tests, the Friedman and the Wilcoxon signed ranks tests, instead. As for DALI, an outlier in the baseline condition was removed and changed to the next highest value + 1.

For SUS, the score for each participant was calculated by multiplying the sum of each item in the questionnaire by a factor of 1.67 to get a score between 0-100. A grand mean was then calculated based on all participants' SUS-scores. For SASSI, negatively phrased questions were reversed, so that higher ratings represented better scores for all questions. A total score and a mean for each factor was calculated for each participant. Then, a grand mean was calculated based on each participants' total score and factor means. As for the DALI questionnaire, a raw DALI score was calculated by summarising all questionnaire items for each participant. Then, a grand mean including data from all participants was calculated.

A thematic analysis was conducted to analyse the interview data. As the interviews were conducted to complement data from the questionnaires, the themes that were used to categorise the data were the factors in the SASSI questionnaire: System Response Accuracy, Likeability, Cognitive Demand, Annoyance, Habitability and Speed. An additional theme, functionality, was added.

## 4 Results

This section presents the results from the analysis.

### 4.1 Cognitive load

Cognitive load was measured using both the DRT and the DALI questionnaire. The results are presented below.

#### Response time

Table 1: The mean response time and hit rate for the baseline (B) condition and the two interfaces, speech (S) and visual-manual (VM), for each participant.

Participant	B (RT)	S (RT)	VM (RT)	B (HR)	S (HR)	VM (HR)
4	506.42	791.6	778.82	1	0.52	0.38
5	361.1	466.85	682.05	0.97	0.94	0.87
6	337	564.49	846.69	1	0.85	0.69
7	269.61	465.02	684.34	1	0.93	0.77
8	407.38	898.18	995.95	1	0.88	0.88
9	362.39	444.54	601.76	1	0.97	0.77
10	624.59	634.61	943.16	0.97	0.91	0.88
11	376.89	709.24	1001.43	0.97	0.74	0.50
12	326.5	491.2	713.39	0.98	0.88	0.85
13	478.73	747.55	960.15	1	0.8	0.56
14	625.59	987.34	997.76	0.85	0.71	0.62
<b>TOTAL</b>	425.11	654.60	836.86	0.98	0.83	0.71

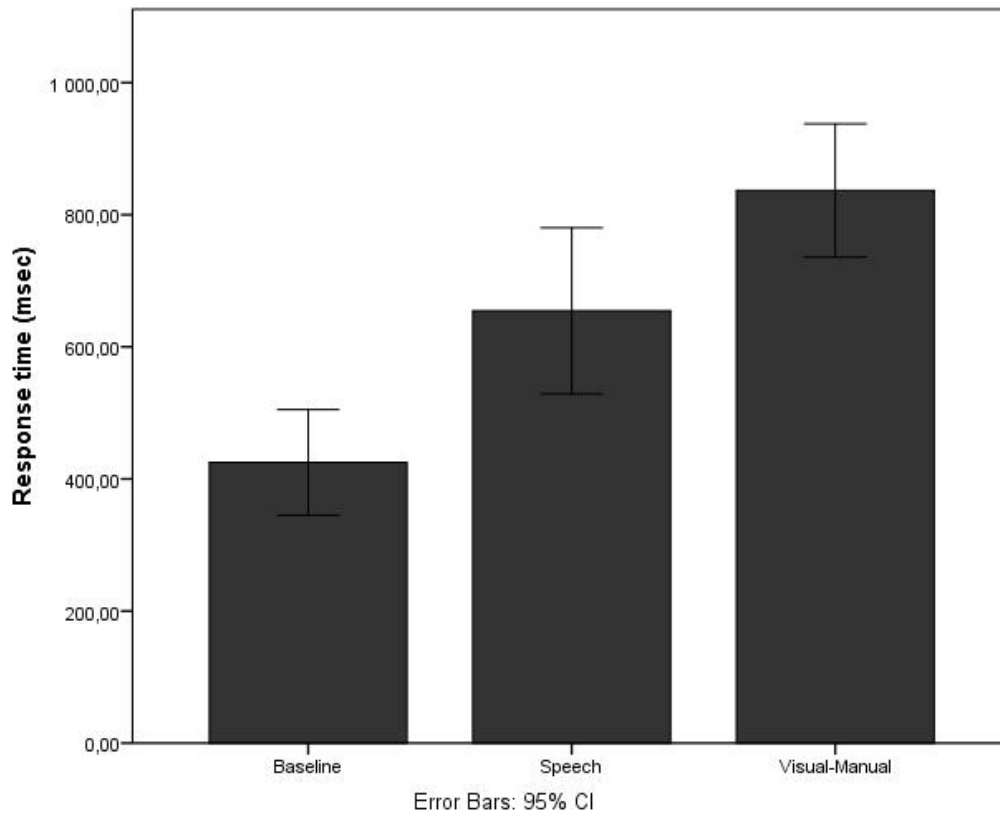


Figure 1: The response time for the three different conditions.

A repeated-measures ANOVA for the three conditions was conducted to compare the effect baseline, speech interface and visual-manual interface had on response time, which showed that type of interface had a significant effect on overall response time,  $F(2, 20) = 59.89$ ,  $p < .001$ ,  $\eta_p^2 = .86$ . Mauchly's test indicated that the assumption of Sphericity was met,  $\chi^2(2) = .69$ ,  $p > .05$ . Post hoc analysis with a Bonferroni adjustment revealed that the overall response time for the baseline ( $M=425.11$ ,  $SD=119.08$ ) was statistically significantly lower than for both the speech interface ( $M=654.60$ ,  $SD=187.34$ ) ( $p=.001$ ) and the visual-manual interface ( $M=836.86$ ,  $SD=150.33$ ) ( $p < .001$ ). The post hoc analysis did also reveal that the speech interface had a statistically significantly lower response time than the visual-manual interface ( $p=.001$ ). The analysis was repeated for each of the three tasks, telephone (task 1), music + note a reminder (task 2) and navigation (task 3).

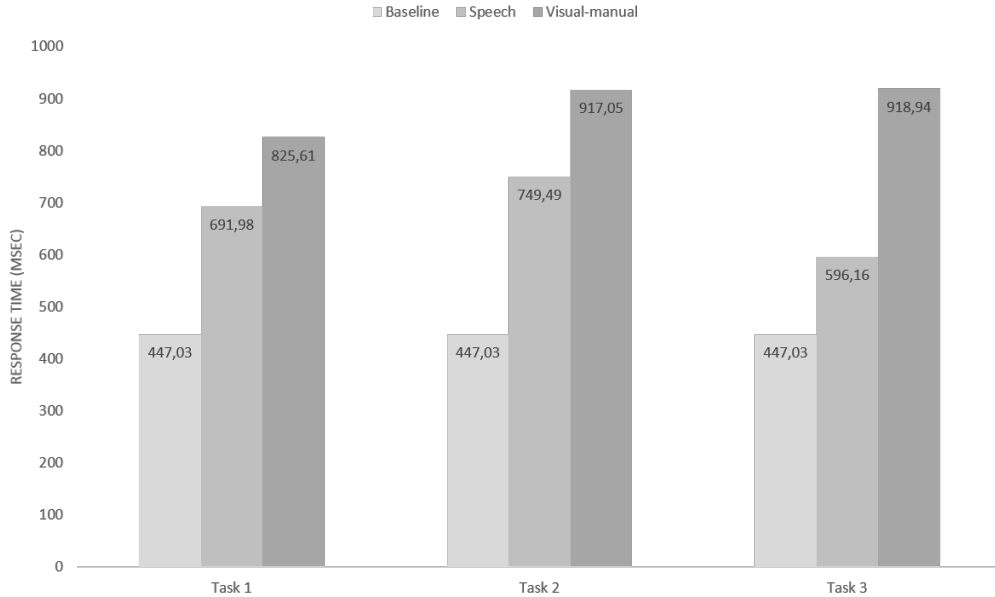


Figure 2: Response time for the three different conditions per task.

**Task 1.** Mauchly’s test of Sphericity indicated that the assumption of sphericity was met,  $x^2(2) = 3.32$ ,  $p > .05$ . A one way repeated-measures ANOVA showed a significant main effect,  $F(2, 20) = 19.61$ ,  $p < .001$ ,  $\eta_p^2 = .66$ . A post hoc test with the Bonferroni correction revealed that baseline had a significantly lower response time than for both the speech ( $p < .001$ ) and the visual-manual interface ( $p = .001$ ). However, the speech interface did not have a significantly lower response time than the visual-manual interface ( $p = .282$ ).

**Task 2.** As for response time for task 2, the Mauchly’s test indicated that the assumption of sphericity had been met,  $x^2(2) = 3.28$ ,  $p > .05$ . The results showed that the response time was significantly affected by the type of interfaces used,  $F(2, 20) = 18.27$ ,  $p < .001$ ,  $\eta_p^2 = .65$ . A post hoc test using Bonferroni showed that the response time for baseline was statistically significantly lower than for speech ( $p = .002$ ) and visual-manual ( $p < .001$ ). Response time for speech was not statistically significantly lower than for the visual-manual task ( $p = .940$ ).

**Task 3.** Mauchly’s test of sphericity showed that the assumption was met,  $x^2(2) = .24$ ,  $p > .05$ . The main effect revealed that there was a sig-

nificant effect of interface on response time,  $F(2, 20) = 40.14$ ,  $p < 0.001$ ,  $\eta_p^2 = .80$ . A post hoc test using pairwise comparisons with the Bonferroni correction revealed that the baseline ( $M = 425.11$ ,  $SD = 119.08$ ) was statistically significantly lower than both the speech ( $M = 596.16$ ,  $SD = 175.95$ ) and visual-manual interface ( $M = 918.94$ ,  $SD = 168.84$ ). Response time for the speech interface was statistically significantly lower than for the visual-manual interface ( $p = .045$ ).

### Hit rate

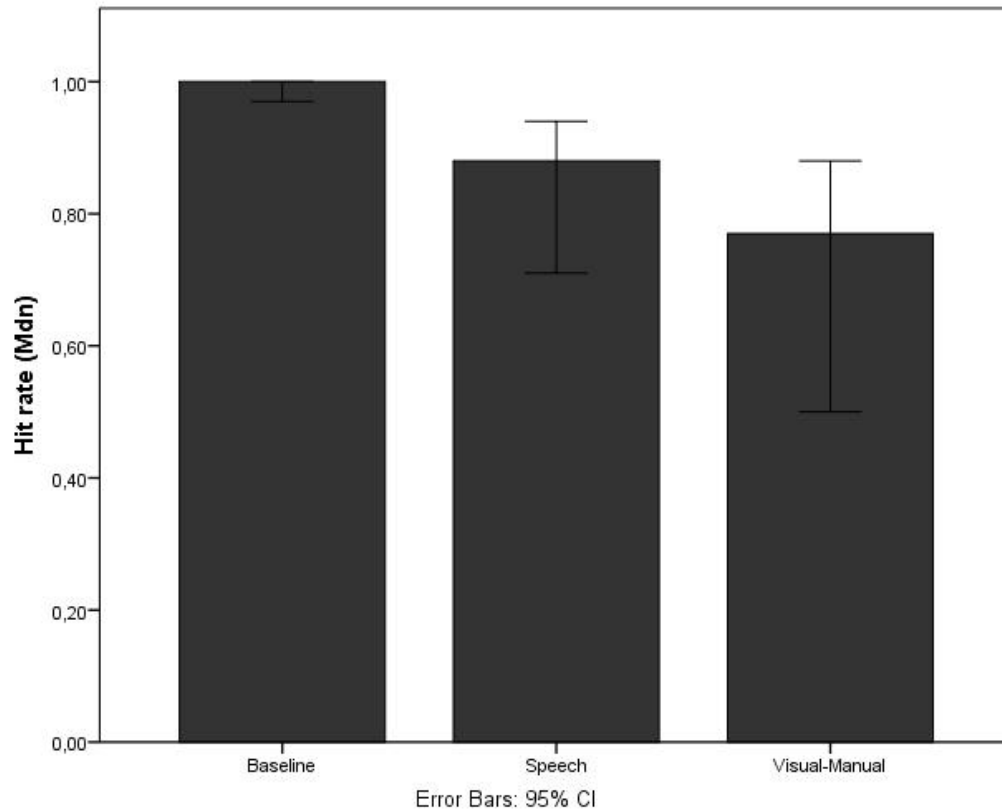


Figure 3: Overall hit rate for the three different conditions.

The Friedman test revealed that there was a statistically significant difference in overall hit rate for the conditions,  $\chi^2(2) = 21.54$ ,  $p < .001$ . A post hoc test in form of the Wilcoxon signed-rank test was conducted with a

Bonferroni correction, so that the significance level was set as  $p < .017$ . All comparisons were statistically significant, baseline (Mnd = 1.00, IQR = 0.03) had a statistically significantly higher hit rate than both speech (Mnd = 0.88, IQR = 0.19) ( $z = -2.936$ ,  $p = .003$ ,  $r = -0.55$ ) and visual-manual interaction (Mnd = 0.77, IQR = 0.33) ( $z = -2.94$ ,  $p = .003$ ,  $r = -0.56$ ). Hit rate for speech was statistically significantly higher than for the visual-manual interface ( $z = -2.81$ ,  $p = .005$ ,  $r = -0.53$ ).

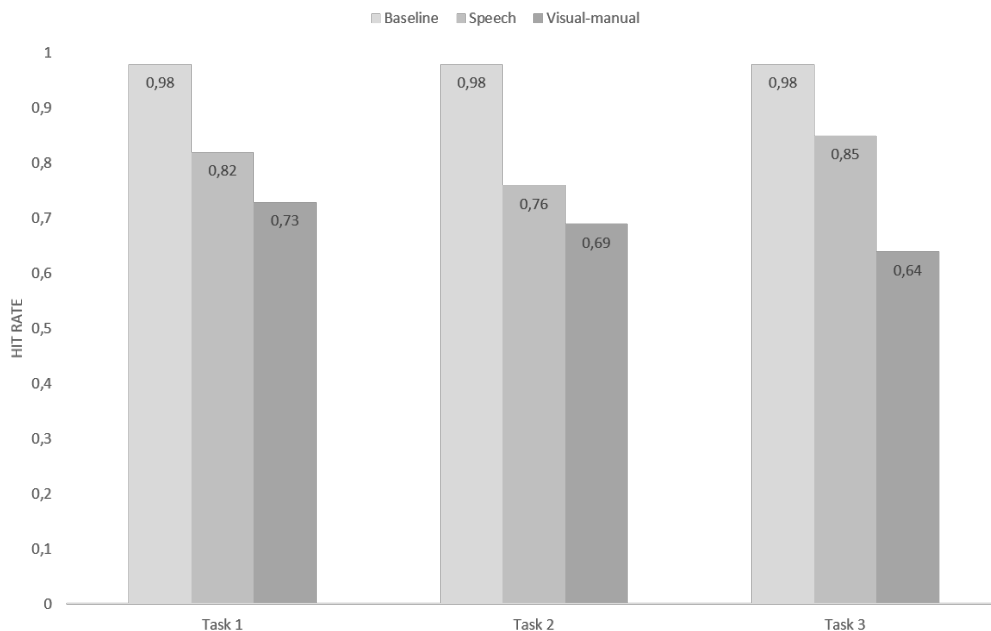


Figure 4: Hit rate for the three different conditions per task.

**Task 1.** The Friedman test revealed an overall main effect of condition on hit rate for task 1,  $x^2(2) = 11.023$ ,  $p = .004$ . A post hoc test using Wilcoxon signed ranks test revealed that baseline (Mnd = 1.00, IQR = 0.03) had a statistically significantly higher hit rate than the speech (Mnd = 0.88, IQR = 0.19) ( $z = -2.93$ ,  $p = .003$ ,  $r = -0.55$ ) and visual-manual interface (Mnd = 0.76, IQR = 0.37) ( $z = -2.93$ ,  $p = .003$ ,  $r = -0.55$ ). Speech did not have a statistically significantly higher hit rate than the visual-manual hit rate ( $p > .017$ ).

**Task 2.** The Friedman test revealed an overall main effect of condition on hit rate for task 2,  $x^2(2) = 16.79$ ,  $p < .001$ . A post hoc test using Wilcoxon

signed ranks test revealed that baseline (Mnd = 1.00, IQR = 0.03) had a statistically significantly higher hit rate than both speech (Mnd = 0.83, IQR = 0.24) ( $z = -2.81$ ,  $p = .005$ ,  $r = .59$ ) and visual-manual (Mnd = 0.78, IQR = 0.37) ( $z = -2.94$ ,  $p = .003$ ,  $r = -0.63$ ). Hit rate for speech was not statistically significantly higher than the visual-manual hit rate ( $p > .017$ ).

**Task 3.** The Friedman test revealed an overall effect of task on hit rate,  $\chi^2(2) = 15.85$ ,  $p < .001$ . Post hoc test using Wilcoxon signed ranks test revealed that hit rate for baseline (Mnd = 1.00, IQR = 0.03) was statistically significantly higher ( $p = .003$ ) than for the visual-manual interface (Mnd = 0.70, IQR = 0.35). However, not compared to the hit rate for the speech interface (Mnd = 0.86, SD = 0.25) ( $p = .058$ ). The speech task had a statistically significantly higher hit rate than the visual-manual counterpart ( $p = .004$ ).



## Driving Activity Load Index - DALI

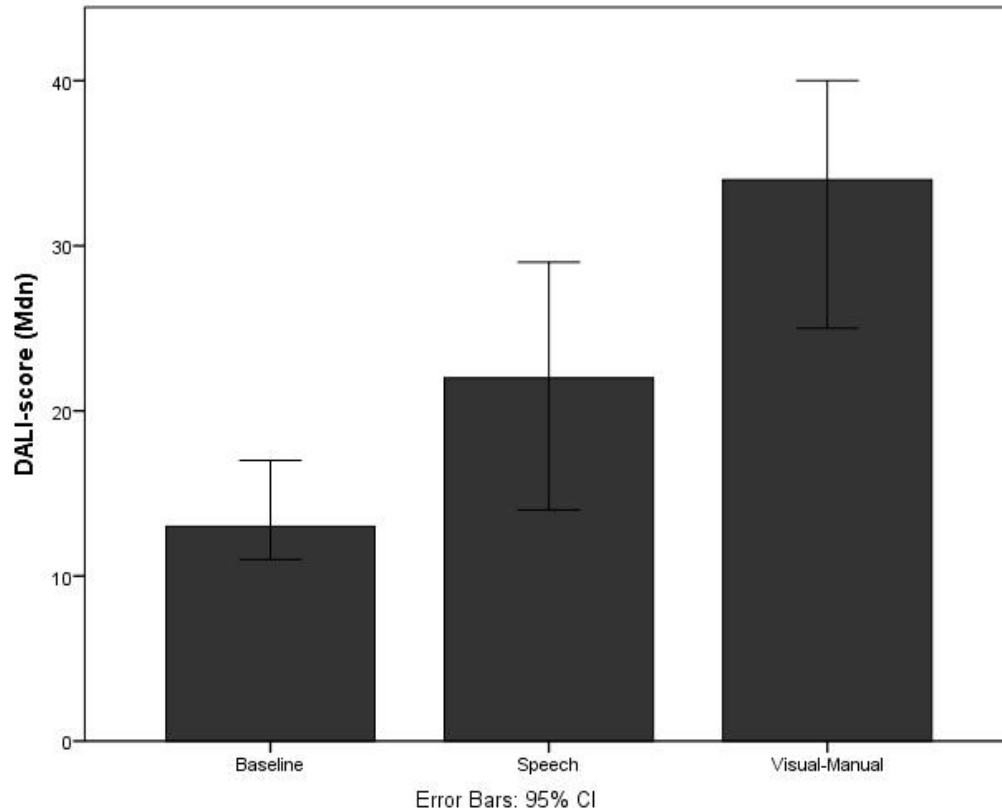


Figure 5: DALI for the three different conditions.

The Friedman test revealed that there was a statistically significant difference in raw DALI-score for the tasks,  $\chi^2(2) = 17.72$ ,  $p < .001$ . A Post-hoc test Wilcoxon signed-rank test was conducted with a Bonferroni correction, so that the significance level was set as  $p < .017$ . Median (IQR) levels for DALI-score for baseline, speech and visual-manual condition were 13 (4), 22 (11) and 34 (9). There was a statistically significant effect between all conditions. DALI-score for baseline was statistically significantly lower than for both speech ( $z=-2.65$ ,  $p = .008$ ,  $r = -0.57$ ) and visual-manual ( $z=-2.94$ ,  $p = .003$ ,  $r = -0.63$ ). DALI-score for speech was statistically significantly lower than for the visual-manual interface ( $z = -2.85$ ,  $p = .004$ ,  $r = -0.61$ ).

## 4.2 Visual distraction

In this section, the results from the eye tracking will be presented. This includes the following measurements: fixation on road, number of glances on task displays and number of glances on task displays greater than 2.0 seconds.

### Fixation on road

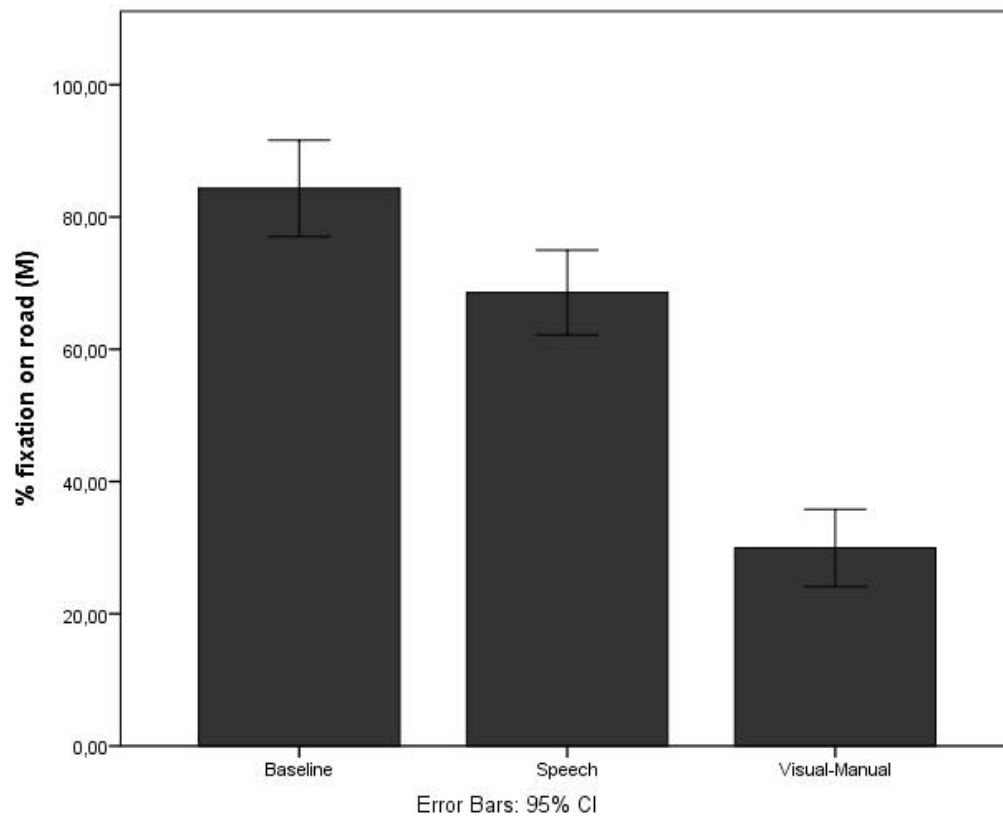


Figure 6: Fixation on road for the three different conditions.

The Friedman test revealed an overall effect of condition on how many percent the drivers fixated on the roadway,  $\chi^2(2) = 16.22$ ,  $p < .001$ . A post hoc test using Wilcoxon signed ranks test revealed a statistically significant difference between all conditions. Baseline (Mdn = 88.98, IQR = 15.47) had

a higher percent of fixation than both the speech (Mnd = 69.52, IQR = 13.89), ( $z = -2.43$ ,  $p = .015$ ,  $r = -0.61$ ) and visual-manual interface (Mnd = 28.37, IQR = 11.43), ( $z = -2.67$ ,  $p = .008$ ,  $r = -0.67$ ). Amount of fixation on road for the speech interface was higher than for the visual-manual interface ( $z = -2.67$ ,  $p = .008$ ,  $r = -0.67$ ).

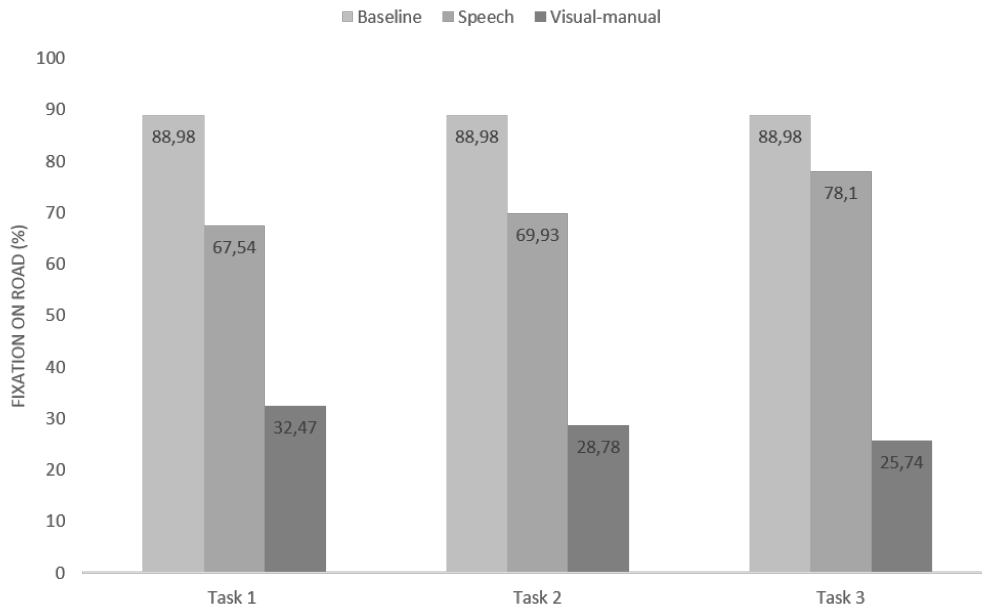


Figure 7: Fixation on road (%) for the three different conditions divided per task.

**Task 1.** The Friedman task revealed an overall effect of condition on fixation,  $\chi^2(2) = 16.22$ ,  $p < .001$ . A post hoc test using the Bonferroni correction revealed that the fixation on road was higher for baseline (Mnd = 88.98, IQR = 15.47) than for the speech (Mnd = 67.54, IQR = 9.15) ( $z = -2.43$ ,  $p = .015$ ,  $r = -0.61$ ) and the visual-manual interface (Mnd = 32.47, IQR = 10.46), ( $z = -2.67$ ,  $p = .008$ ,  $r = -0.67$ ). Comparing fixation on road for the speech and visual-manual interface revealed that the fixation on road was statistically significantly higher for the speech interface ( $z = -2.67$ ,  $p = .008$ ,  $r = -0.67$ ).

**Task 2.** The Friedman test revealed an overall effect of condition on amount of fixation on the road,  $\chi^2(2) = 14.89$ ,  $p = .001$ . A post hoc test using the Wilcoxon signed rank test with a Bonferroni correction revealed

a statistically significant difference between the baseline and visual-manual task ( $z = -2.67$ ,  $p = .008$ ,  $r = -0.67$ ). The difference between the baseline and the speech interface was not statistically significant ( $p > .017$ ). Speech had a statistically significantly higher percent of fixation on the roadway compared to the visual-manual interface ( $z = -2.67$ ,  $p = .008$ ,  $r = -0.67$ ).

**Task 3.** The Friedman test revealed an overall effect of condition on amount of fixation on road,  $\chi^2(2) = 14.89$ ,  $p = .001$ . A post hoc test using a Bonferroni correction revealed that baseline had statistically significantly higher amount of fixations than the visual-manual interface ( $z = -2.67$ ,  $p = .008$ ), but not compared to the speech interface ( $p > .017$ ). Speech had a statistically significantly higher percent of fixations on road compared to the visual-manual interface ( $z = -2.67$ ,  $p = .008$ ,  $r = -0.67$ ).

## Glances on task displays

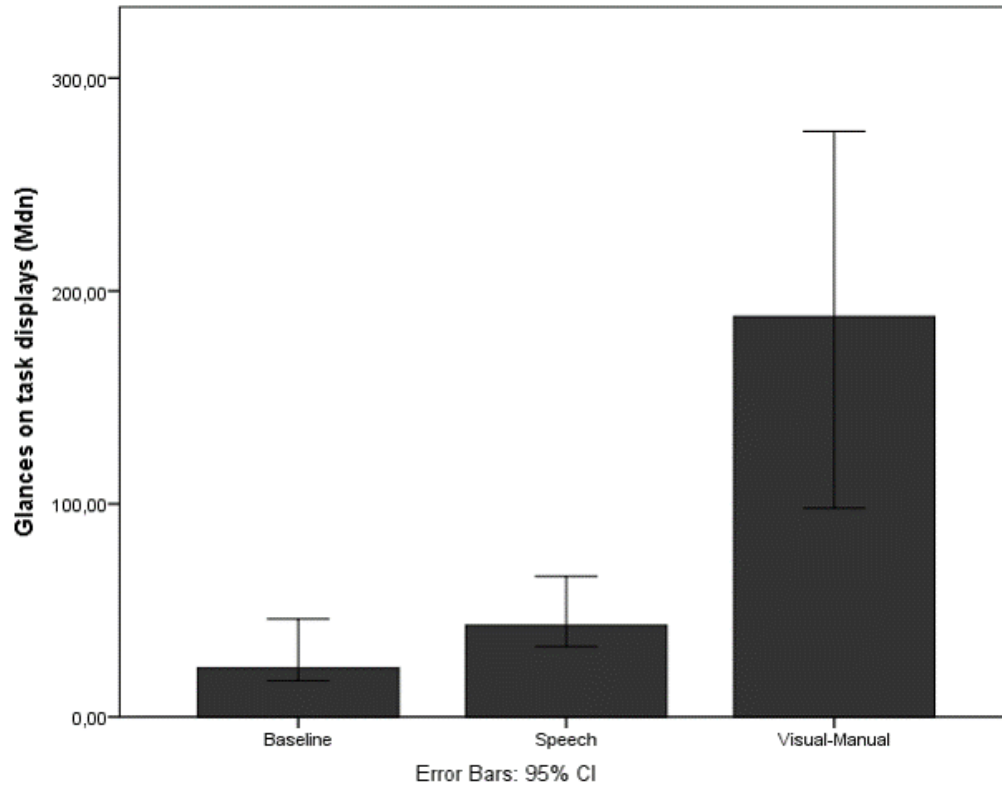


Figure 8: Glances on task displays for the three different conditions.

An overall effect of task for number of glances on task displays was found with the Friedman test,  $\chi^2(2) = 14.89$ ,  $p = .001$ . A post hoc test using the Wilcoxon signed ranks test revealed a significant effect between the baseline (Mnd = 23.00, IQR = 18.00) and visual-manual interface (Mnd = 28.37, IQR = 11.43), ( $z = -2.67$ ,  $p = .008$ ,  $r = -.57$ ). This was not found between baseline and the speech interface ( $p > .017$ ). However, the speech interface (Mnd = 69.53, IQR = 13.89) was found to have a significantly lower amount of glances on task displays compared to the visual-manual counterpart ( $z = -2.67$ ,  $p = .008$ ,  $r = -.57$ ).

## Per task

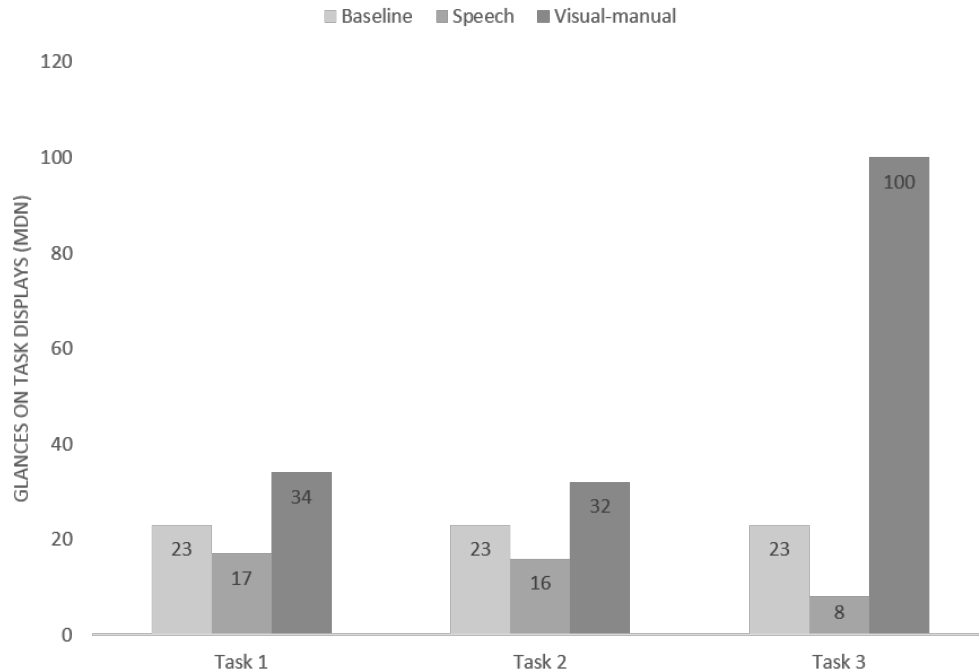


Figure 9: Glances on task displays for the three different conditions per task.

**Task 1.** There was no significant effect for task 1 ( $p > .017$ ).

**Task 2.** The Friedman test revealed an overall statistically significant effect of condition on number of glances on task displays,  $\chi^2(2) = 6.22$ ,  $p = .045$ . The post hoc test revealed that the speech interface (Mnd = 16.00, IQR = 17.50) had a statistically significantly lower amount of glances on task displays compared to the visual-manual counterpart (Mnd = 32.00, IQR = 18.00) ( $z = -2.43$ ,  $p = .015$ ,  $r = -0.61$ ). The difference between the baseline (Mnd = 23.00, IQR = 18.00) and the speech interface was none significant ( $p > .017$ ).

**Task 3.** The Friedman test found an overall significant effect of condition on number of glances on task displays,  $\chi^2(2) = 16.22$ ,  $p < .001$ . A post hoc test with the Wilcoxon signed rank test revealed that baseline (Mnd = 23.00, IQR = 18.00) had a significantly higher amount of glances on task displays compared to the speech interface (Mnd = 8.00, IQR = 10.00), ( $z =$

-2.67,  $p = .008$ ,  $r = -0.67$ ). It was also found that the speech interface had a statistically significantly lower amount of glances than the visual-manual counterpart (Mnd = 100.00, IQR = 114.00), ( $z = -2.67$ ,  $p = .008$ ,  $r = -0.67$ ).

## Glances on task displays with a duration of 2 sec or more

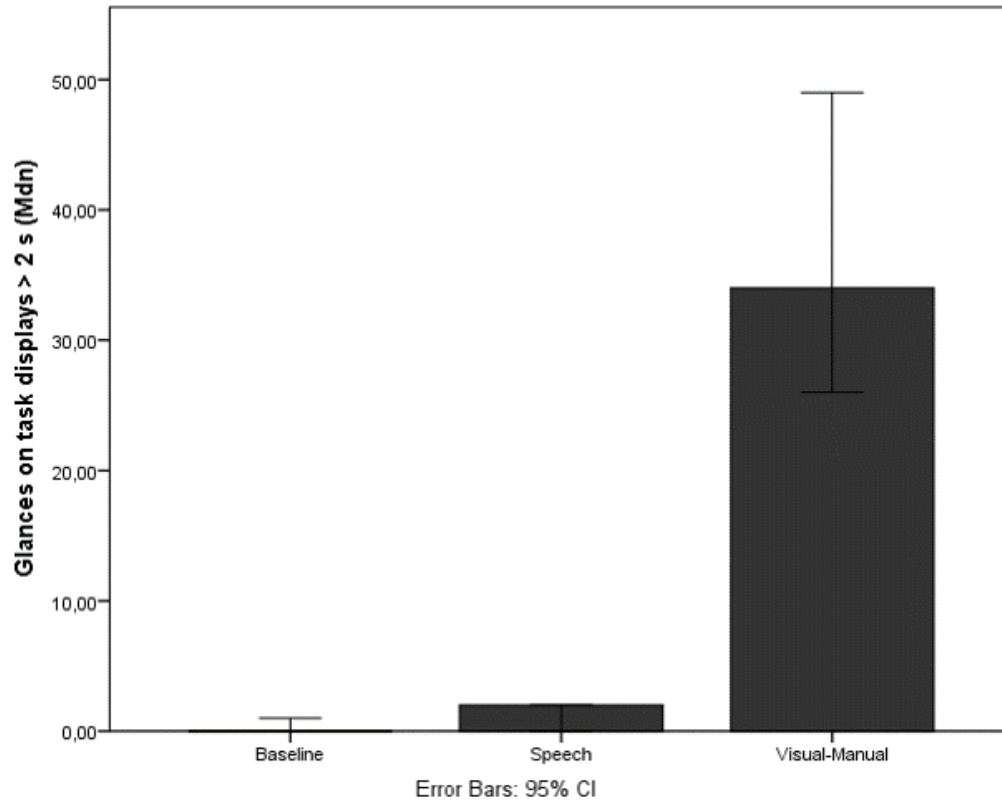


Figure 10: Glances on task displays with a duration of 2 sec or more for the three different conditions.

The Friedman test found an overall effect of condition on glances on task displays with a duration of 2 seconds or more,  $\chi^2(2) = 14.82$ ,  $p = .001$ . Post hoc test with the Wilcoxon signed ranks test revealed a statistically significant difference between the baseline (Mdn = 0.0, IQR = 0.5) and visual-manual interface (Mdn = 34.00, IQR = 21.50), ( $z = -2.67$ ,  $p = .008$ ,  $r = -.57$ ). The speech interface (Mdn = 2.00, IQR = 2.00) had a statistically significantly lower amount of eye glances with a duration of 2 seconds or more compared to the visual-manual interface (Mdn = 34.00, IQR = 21.50), ( $z = -2.67$ ,  $p = .008$ ,  $r = -.57$ ).



## Per task

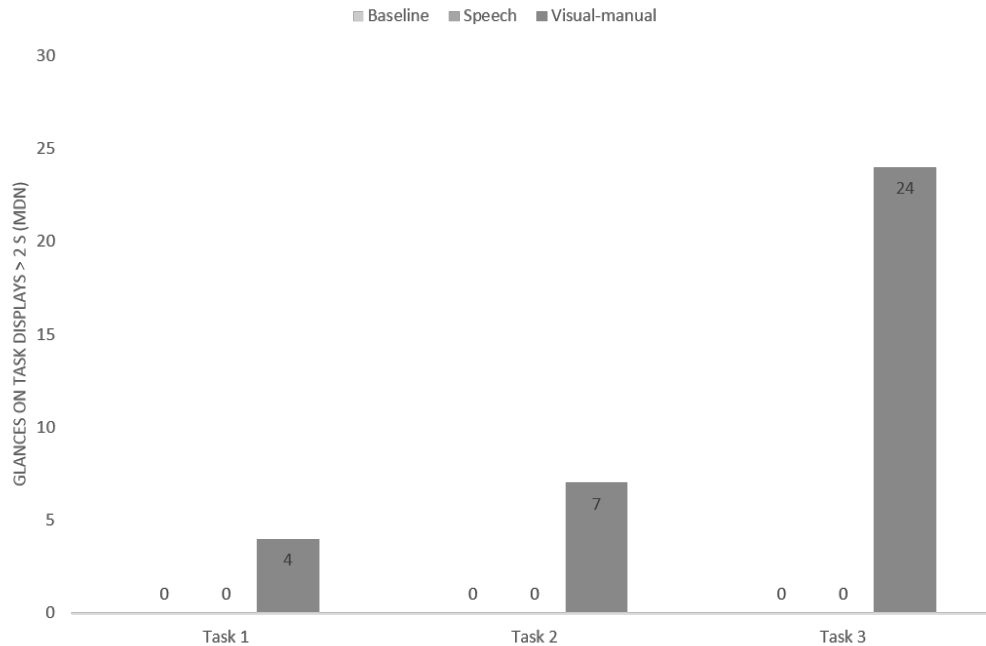


Figure 11: Glances on task displays with a duration of 2 sec or more for the three different conditions per task.

**Task 1.** The Friedman test revealed an overall significant effect of condition on number of glances with a duration of 2 seconds or more,  $\chi^2(2) = 14.00$ ,  $p = .001$ . A post hoc test with the Wilcoxon signed rank test revealed that there was a significant difference between the baseline (Mnd = 0, IQR = 0.50) and visual-manual interface (Mnd = 4.00, IQR = 6.50) ( $z = -2.54$ ,  $p = .011$ ,  $r = -0.64$ ) as well as between the speech (Mnd = 0, IQR = 1.00) and visual-manual interface ( $z = -2.67$ ,  $p = .008$ ,  $r = -0.67$ ). There was no significant difference between the baseline and speech interface ( $p > .017$ ).

**Task 2.** The Friedman test revealed an overall significant effect,  $\chi^2(2) = 13.94$ ,  $p = .001$ . A post hoc test with the Wilcoxon signed rank test found that baseline had a significantly lower number of glances with a duration of 2 seconds or more compared to the visual-manual interface (Mnd = 7.00, IQR = 4.50), ( $z = -2.67$ ,  $p = .008$ ,  $r = -0.67$ ). A significant effect was also found between the speech (Mnd = 0, IQR = 1.00) and visual-manual interface

( $z = -2.52$ ,  $p = .012$ ,  $r = -0.63$ ), showing that the speech interface had a significantly lower number of glances than the visual-manual task.

**Task 3.** The Friedman test revealed an overall effect  $\chi^2(2) = 15.25$ ,  $p < .001$ . Baseline (Mnd = 0, IQR = 0.50) had statistically significantly lower amount of glances with a duration of 2 seconds or more than compared to the visual-manual task (Mnd = 24.00, IQR = 21.00), ( $z = -2.67$ ,  $p = .008$ ,  $r = -0.67$ ). Speech (Mnd = 0, IQR = 1.00) had a significantly lower number of glances with a duration of 2 seconds or more than compared to the visual-manual interface ( $z = -2.67$ ,  $p = .008$ ,  $r = -0.67$ ). There was no significant effect between the baseline and speech interface ( $p = 1.00$ ).

### 4.3 User acceptance and efficiency

The results from SUS, SASSI and the interviews are presented below.

#### System Usability Scale - SUS

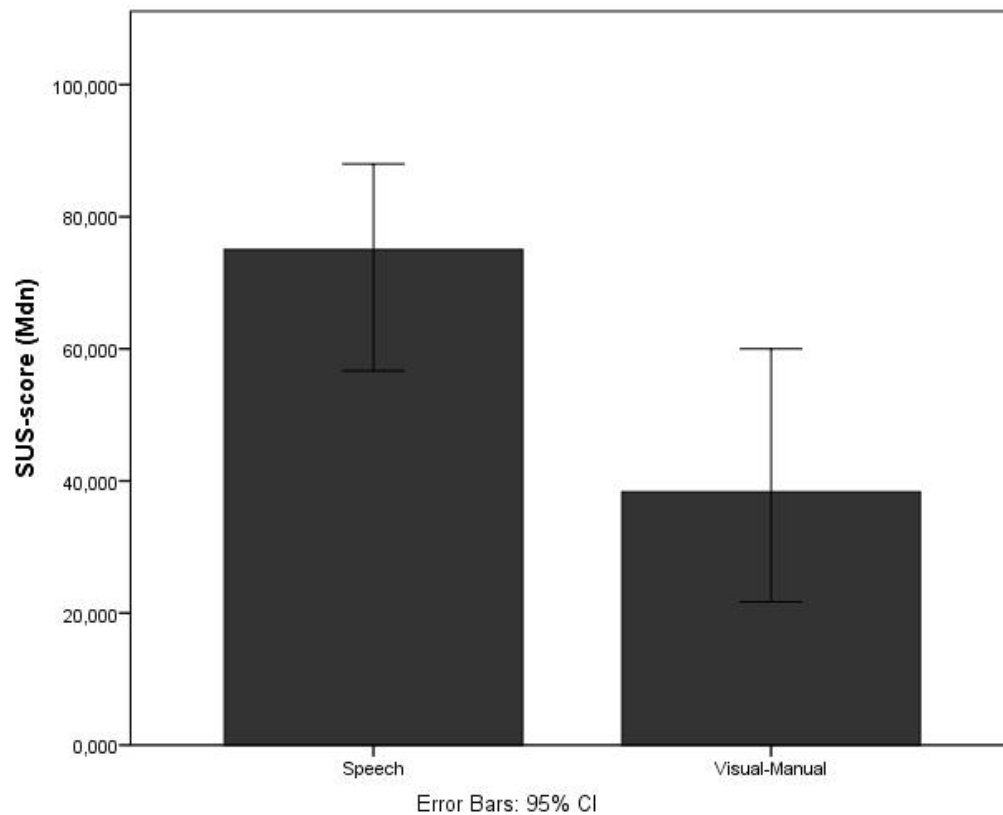


Figure 12: SUS-score for the two interfaces.

SUS-score for speech is statistically significantly higher for the speech interface (Mdn = 75.02, IQR = 18.86) compared to the visual-manual interface (Mdn = 38.34, IQR = 21.67), ( $z = -2.85$ ,  $p = .004$ ,  $r = -0.61$ ).

## Subjective Assessment of Speech Interfaces - SASSI

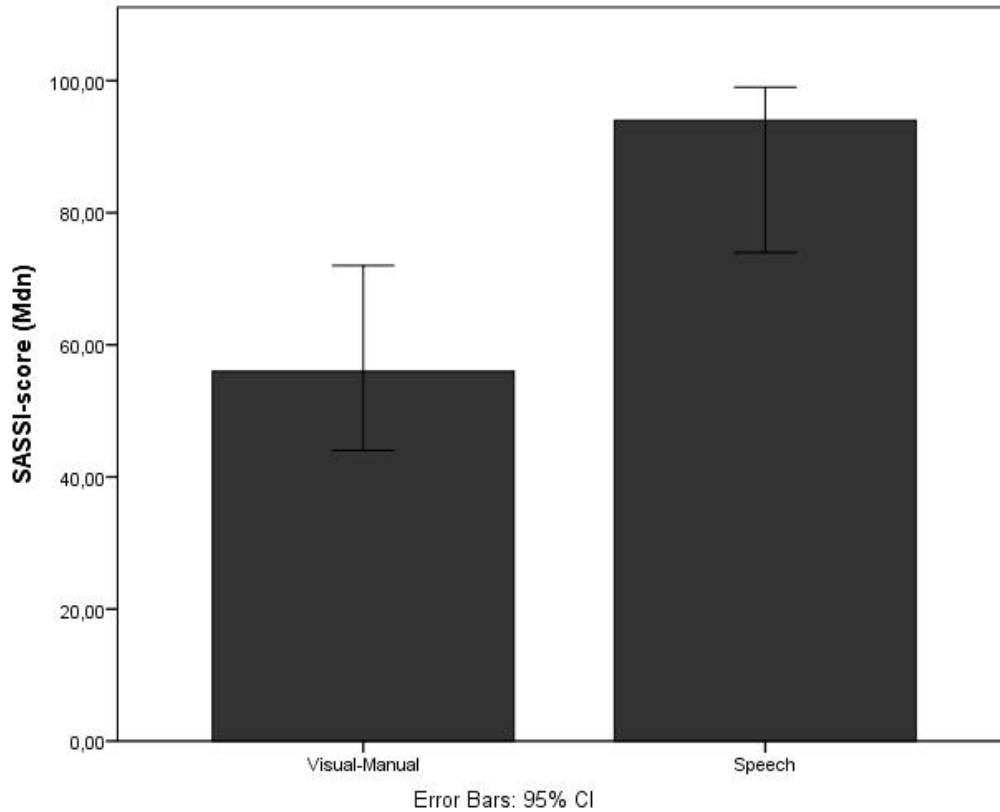


Figure 13: SASSI-score for the two interfaces.

Wilcoxon signed ranks test revealed that the overall SASSI-score for the speech interface (Mnd = 94.00, IQR = 17.00) had a statistically significantly higher score than the visual-manual interface (Mnd = 56.00, IQR = 14.00), ( $z = -2.85$ ,  $p = .004$ ,  $r = -0.61$ ). Comparing the factors did also reveal a result in favour of the speech interface in regards to cognitive demand, annoyance, habitability and likeability. *Cognitive demand* (Mnd = 5.20, IQR = 1.40) was significantly lower for the speech interface than compared to the visual-manual interface (Mnd = 3.00, IQR = 1.40), ( $z = -2.81$ ,  $p = .005$ ,  $r = -0.59$ ). Also *annoyance* was statistically significantly lower for speech than for the visual-manual interface ( $z = -2.66$ ,  $p = .008$ ,  $r = -0.57$ ). *Habitability* had a significantly higher value for the speech interface (Mnd = 4.67, IQR = 1.00)

than for the visual-manual interface (Mnd = 2.67, IQR = 1.33), ( $z = -2.40$ ,  $p = .016$ ,  $r = -0.52$ ). The factor *likeability* was statistically significantly higher for the speech interface (Mnd = 5.80, IQR = 1.20) than for the visual-manual interface (Mnd = 4.00, IQR = 1.40), ( $z = -2.85$ ,  $p = .004$ ,  $r = -0.61$ ).

## **Interviews**

This sections presents the themes and results from the interviews.

### **System Response Accuracy**

Four of the participant's explicitly said that they felt that the speech system did not understand what they were saying. Five mentioned that the system understood them sometimes, but that it sometimes did not. They had to repeat the commandos and modify them, such as adding 'thanks' after 'yes' because the system had problems understanding short commandos. Two of the participants felt that they were always understood by the system. None of the truck drivers mentioned being non-satisfied with the response from the speech system.

### **Likeability**

The speech system was mentioned as being easy to learn by two participants. Another participant described it as being very simple. One participant said that he never uses speech system, but would if it would be as easy to use as this one. Two participants mentioned it being better than other speech systems tested.

### **Cognitive demand**

Seven of the drivers explicitly said that interacting with the speech system enabled them keeping their attention on the roadway.

### **Annoyance**

Four of the drivers felt that they had to repeat the commandos several times in order for the system to understand what they were saying, which lead to annoyance. One participant felt stressed using the system while driving.

### **Habitability**

Five of the drivers said that they did not know what to say to the system, but that they thought it would be easy to learn. Three of the participants felt

that they always knew what to say to the system. One participant felt that for some commandos he knew, but that it was harder for other commandos. Two participants felt that they did not know what to say to the system. One participant thought that the commando for getting resting times, 'When do I have my break?' should be changed to 'When do I need to take a break?' as this is more specific.

### **Speed**

Three of the participant mentioned that the system was slow and took time to process the commandos given.

### **Functionality**

The following are voice functionality that the truck drivers suggested would be useful to add:

- Adjust windows
- Dryers
- Sunroof
- Lights
- Open/close the door
- Adjust seat
- Horn
- Set alarm when sleeping in truck
- Communicate with the truck through phone, such as the engine heater
- Dynafleet
- Climate control
- Ask about weights and axle load
- Change display appearance (get black panel)

- Full and dimmed lights

All suggestions of functionality were mentioned once, except for controlling Dynafleet which two truck drivers mentioned. However, many of the truck drivers felt that the functionality implemented was enough.

## 4.4 Summary

This section summaries the most important findings in the result section.

### Cognitive load

The objective measurement of cognitive load, TDRT, revealed a result of the speech interface being less demanding than the visual-manual interface when looking at data from all tasks. Further, it was found that the truck drivers perceived the speech interface as less demanding than the visual-manual interface which was shown by the DALI questionnaire.

An analysis was made looking into data from each task. This showed that the largest effect was for the navigation task, which was shown to have a significantly lower response time and higher hit rate for speech compared to the visual-manual interface. For hit rate, baseline was significantly lower than the speech tasks for all tasks except for the navigation task. For the navigation task, the results did not reveal any significant difference between baseline and the speech interface.

### Visual distraction

Measurement for the eye glance data were fixation on road (%), number of glances on tasks displays and number of glances on task displays with a duration of 2 seconds or more. All of these measurements showed a result in favour of the speech system. Overall *fixation on road* revealed that the speech interface enables the drivers to fixate more on the road compared to the visual-manual interface. The baseline condition had a higher fixation on road than both interfaces. For all separate tasks, speech had a significantly higher percent fixation on road than the visual-manual counterpart. When looking into *number of glances on task displays* the different interfaces required, the overall data revealed that the speech interface required fewer glances than the visual-manual counterpart. Looking into each task, both task 2 and

3 showed that the speech interface required fewer glances than the visual-manual counterpart. Especially the navigation task stood out with median glances of 100 for the visual-manual interface, 8 for the speech interface and 23 for the baseline. The results show that the speech interface had a significantly lower amount of glances on task displays even compared to the baseline.

For *glances with a duration of 2 seconds or more*, the overall data showed that the speech interface requires a lower number of glances than the visual-manual counterpart. This was also the case for all tasks.

### **User acceptance and efficiency**

The SUS-score was higher for speech (Mnd = 75.02, IQR = 18.86) than for the visual-manual interface (Mnd = 38.34, IQR = 21.67). The result for the speech interface was above the average score, but this was not the case for the visual-manual interface. Also SASSI was higher for speech (Mnd = 94.00, IQR = 17.00) than for the visual-manual interface (56.00, IQR = 14.00). Further, the factors cognitive demand, annoyance, habitability and likeability all revealed a result in favour of the speech system.

The interviews highlighted some issues and areas to work on for the speech system. Some participants felt that the system did not understand them leading to them having to repeat commandos, some participants did not know what to say to the system and some thought the system was slow. The interviews also revealed some functionality that can be added to future versions of the system. Further, positive aspects were highlighted such as the system was perceived as being easy to learn and simple. Many truck drivers felt that it enabled them having their attention on the roadway and that they would use the system if they had it available.



## 5 Discussion

This section will first present and discuss the results divided into each research question and then a discussion about the methods used.

### 5.1 Results

#### **How does the speech interface as compared to the visual manual interface affect distraction while driving?**

First, the results for distraction deriving from cognitive load will be discussed. The first analysis was made looking into overall data including all tasks. This revealed that the speech interface is less distracting to use while driving compared to the visual-manual interface. The subjective measurement of cognitive load strengthens the results of the speech interface being less distracting by showing that the truck drivers also *perceived* the speech system as being less demanding while driving than compared to the visual-manual interface. Further, the results from the interviews show that many of the truck drivers explicitly expressed that they felt the speech system was letting them have their attention on the roadway while they were driving.

The next analysis was made looking into each task separate. This revealed that the objective measurement of cognitive load found the navigation task to be less demanding when using the speech interface compared to the visual-manual interface. This was also the task where the results for the speech interface was closest to the results for the baseline tasks in regards to hit rate data. However, the calling task, playing a song and noting a reminder did not reveal any significant results in the comparison between the interfaces. When looking into mean response times, the visual-manual interface have 825.61, 917.05 and 918.94 for the three tasks respectively. In comparison, the speech interface has 691.98, 749.49 and 596.16. The results show that the response time for the tasks in the visual-manual interface are quite similar when comparing between the tasks, but that it differs more for the tasks within the speech interface. The response time are higher for task 1 and 2 and lower for task 3, which is the navigation task. Lee et al. (2001) showed that more complex speech tasks impose a higher demand on the drivers meaning that different interfaces and tasks can be more or less distracting depending on the complexity. For this particular speech interface the navigation task seem to be the less complex tasks when done through the speech interface.

The results for visual distraction show that the speech interface is less visually distracting than the visual-manual interface for all measurements and tasks (except for number of glances for task 1). The speech interface let the drivers have a higher fixation on the roadway and demands fewer glances on tasks displays. Important to note is that these results are despite the speech system giving visual stimuli to the driver. As mention before, number of glances on tasks displays for the calling task was the only task where the visual-manual counterpart was not requiring significantly more glances than the speech system. The reason could be that the drivers are so used to carrying out the task that it is almost automated.

### **Is interaction via voice a safer and more efficient way to communicate compared to traditional interfaces?**

Several researcher have found a strong connection between unsafe driving behaviour and visual distraction (For example Hickman et al., 2010: Engström, 2011: Victor et al., 2014). This means that if an interface is less visually distracting than the other it should also be the safer alternative. As described in the result and discussion of the research question above, the speech interface is less visually distracting than the visual-manual interface. The measurement with the strongest connection to safety is glances greater than 2.0 seconds which NHTSA (2012) recommends to avoid as it is proven to increase crash risk. For this measurement, the visual-manual interface had a significantly higher number of glances for all tasks, which further supports that the speech interface is safer than the visual-manual counterpart. However, distraction deriving from visual demand can not alone be used to answer the question.

Cognitive load and its connection to safety is not as conclusive as visual distraction. When looking at overall data it is clear that the visual-manual interface is imposing a higher cognitive load than the speech interface. On task level, it is shown that the navigation task carried out with the visual-manual interface is more cognitively demanding than compared to the speech interface. Both interfaces are therefore imposing cognitive load on the truck drivers, but the speech interface does so less than the visual-manual counterpart. With both measurements pointing in a direction of the speech interface being safer than the visual-manual interface, it might be more interesting to further investigate the differences between baseline driving and driving while interacting with the speech interface to see if the speech interface is safe

enough. It seems like the speech interface imposes a higher cognitive load than the baseline task, but in regards to number of glances on tasks displays for the navigation task it was revealed that the speech task (Mnd = 16.00, IQR = 17.50) had fewer glances than compared to the baseline task (Mnd = 23.00, IQR = 18.00). An explanation for this might be that the baseline data was recorded during a longer time and that the glances could reflect the drivers for example adjusting settings in the truck. However, the result could mean that the drivers were more focused during the speech task than compared to when they were just driving, which could mean that using the speech interface results in a safer driving behaviour. In some studies, cognitively demanding tasks have been shown to lead to a safer driving behaviour compared to when just driving (Engström, 2011; Victor et al., 2014). This should be investigated further by adding other measurements such as lane keeping, brake response or measuring response times for road obstacles.

### **What is the truck drivers acceptance and perceived efficiency of the speech interface compared to it's visual-manual counterpart?**

The SUS and SASSI scores both show that the acceptance and perceived efficiency are higher for the speech system than compared to the visual-manual system. For SUS, the median score is 75.02 which is above 70 and therefore is what Bangor et al. (2008) means an acceptable score. The visual-manual interface got a median score of 38.34 which is far from the acceptable score of 70 or above. Using SUS as a measurement clearly reveals a result in favour of the speech system. Also SASSI reveals that the overall score as well as the score for the factors cognitive demand, annoyance, habitability and likeability are better for the speech system compared to the visual-manual counterpart. As for the interviews, especially the theme 'cognitive demand' illustrated that the participants prefer the speech interface before the visual-manual interface thinking that the speech interface was less cognitively demanding than the visual-manual interface.

## **5.2 Method**

Different aspects of the chosen method and the study set-up can influence the results. This will therefore be discussed below.

## **Study set-up**

The pilot study was carried out on an open road, but the real study was conducted at Hällered's proving ground on a motorway track. The drivers were employed test drivers working at the proving ground.

Because of this, the track was well-known for the test drivers which all had driven there several times. Many truck drivers do drive the same road repeatedly, but it is important to note that there are truck drivers that go to entirely new places which might be more demanding and distracting than driving on a road on which you have driven several times. Therefore, the results from this study does not represent using these interfaces while driving in totally new places as the demand probably will increase.

Many other different types of driving conditions exist as well, such as driving in cities, on country roads or in forests. The track used was a motorway, so driving on other types of roads could give other results. Another aspect is the traffic density during the tests. There were other vehicles driving on the track whilst the tests were carried out, but the traffic density was low. Driving when the traffic density is moderate or high could be more demanding and the outcome of carrying out distracting secondary tasks worse.

As mentioned above, the participants in the study were employed as test drivers with the assignment to test the trucks including the in-vehicle systems. Because of the nature of their work, the test drivers might have a higher interest in new technology and as they test the interfaces it often leads to them becoming experts on the existing interfaces in the trucks. However, the knowledge of the visual-manual interface used in the study did differ, which means that they still represent different types of experience levels. But it is important to discuss that there are many different types of truck driver segment which have different needs and goals.

## **Comparing interfaces**

The most important aspect for the study design chosen in this study was that the conditions was approximately the same for the two interfaces, so that an adequate comparison could be made. As the study design was to compare two different interfaces, the task chosen in the interfaces had to be representative of how the truck drivers normally carries out the tasks. The experience and knowledge level for the users also have to be approximately the same in order for a comparison to be fair.

The procedure for choosing the tasks was the following. The functionality in the speech system was first decided and after that the visual-manual counterpart was settled. The visual-manual counterpart was decided based on an observational study made in the initial user study of the project. The calling task, playing music task, navigation task, looking at resting times as well as looking up warning signs did all have a clear counterpart in the existing visual-manual in-vehicle system. However, the 'note' task did not have an obvious counterpart in the visual-manual interface. But according to the observations made many truck drivers used paper and pen to note information down while driving. The truck drivers could for example receive a phone call in which they got new information they had to remember and therefore wrote it down on a piece of paper. Based on this, it was decided to use paper and pen for the visual-manual counterpart of noting down information. Another alternative could have been to use a note application in their smartphone, but carrying out the task in that way was not observed in the user study.

The test drivers all had different levels of experience of using the visual-manual in-vehicle system used in the study and none of the drivers had any experience in using the speech system. Giving the drivers time to train on the task until they felt stable for both interfaces was a way of making sure of a more adequate comparison of the interfaces. However, there will be a difference in experience levels and skills between participants and interfaces which might affect the results.

It is also important to note that the speech system gives visual stimuli to the truck drivers, even though the tasks could be carried out without viewing any of the task displays. It is therefore not a comparison against an interface being totally free from visual demand. A speech system entirely free from visual stimuli could give other results. The results should not be generalised to all speech interfaces or all visual-manual interfaces. The results in this study reveals results for these two specific interfaces.

### **Measurements used**

The use of several measurements as well as using both objective and subjective measurements was a way of triangulating the results. The objective measurements revealed how distracting the interfaces were and the subjective measurement revealed how distracting the participants perceived the interfaces to be. The subjective measurement measured the overall subjective

mental workload for the baseline condition and the two interfaces. Interesting would have been to add a subjective measurement of cognitive load on a task level, as the objective measurement was used both on an overall level as well as on a task level. Using measurement of cognitive load was proven to be a good way of comparing the different conditions and tasks, showing which tasks that were more demanding as well as giving a reference point by collecting baseline data. However, the connection between cognitive load and driving behaviour need to be investigated further.

The questionnaires, SUS, DALI and SASSI, gave an overall estimate of the interfaces and the interviews complemented this by adding specific areas that were sufficient and which areas that were in need of improvement. Other measurements that could have been used are measurements with a clear connection to driving behaviour such as lane keeping or measuring response times for road obstacles.

## 6 Conclusions and further research

A comparison of a speech interface and its visual-manual counterpart in regards to distraction, safety and the user acceptance and perceived efficiency has been conducted. The conclusions were that (1) the speech interface was less visually distracting than the visual-manual counterpart, (2) the speech interface was less cognitively demanding than the visual-manual interface, especially in the navigation task, (3) the speech interface was safer to use while driving compared to the visual-manual interface and (4) the speech interface had higher user acceptance and efficiency than the visual-manual interface.

Further research should investigate the connection between cognitive load and driving behaviour, such as lane keeping and brake response time, by employing a variety of speech tasks with various complexity as well as including speech interfaces entirely free from visual demand. The focus should be on the differences between baseline driving and speech interaction, as opposed to speech interaction and visual-manual interaction, especially as some studies have shown that cognitively demanding tasks results in a safer driving behaviour.

## References

- Ayres, P., & Paas, F. (2012). Cognitive Load Theory: New Directions and Challenges. *Applied Cognitive Psychology*, 26(6), 827-832.  
doi:10.1002/acp.2882
- Baddeley, A. D. (2006). Working memory: An overview. *Working memory and education*, 1-31.
- Bangor, A., Kortum, P. T., & Miller, J. T. (2008). An empirical evaluation of the system usability scale. *Intl. Journal of Human-Computer Interaction*, 24(6), 574-594.
- Bannert, M. (2002). Managing Cognitive Load Recent Trends in Cognitive Load Theory. *Commentary. Learning And Instruction*, 12(1), 139-46.
- Barón, A., & Green, P. (2006). Safety and usability of speech interfaces for in-vehicle tasks while driving: A brief literature review (No. UMTRI-2006-5). University of Michigan, Transportation Research Institute.
- Bergstrom, J. R., & Schall, A. J. (2014). Eye tracking in user experience design. [Elektronisk resurs]. Waltham, MA : Morgan Kaufmann, c2014.
- Brooke, J. (1996). SUS-A quick and dirty usability scale. *Usability evaluation in industry*, 189(194), 4-7.
- Chin, E., Nathan, F., Pauzie, A., Manzano, J., Nodari, E., Cherri, C., ... & Marchitto, M. (2004). Subjective Assessment Methods for Workload. *AIDE Deliverable*, 2(6).
- Engström, J. (2010). The tactile detection task as a method for assessing drivers' cognitive load. *Performance Metrics for Assessing Driver Distraction: The Quest for Improved Road Safety*, 90-103.
- Engström, J. (2011). Understanding attention selection in driving: From limited capacity to adaptive behaviour. Chalmers University of Technology.
- Engström, J., Victor, T., & Markkula, G. (2013). Attention selection and multitasking in everyday driving: A conceptual model.
- Field, A. (2009). *Discovering statistics using SPSS*. Sage.
- Ghaoui, C. (Ed.). (2005). *Encyclopedia of human computer interaction*. IGI Global.
- Goodwin, K. (2009). *Designing for the digital age. [Elektronisk resurs] : how to create human-centered products and services*. Indianapolis, Ind. : Wiley Pub., c2009.



- Greene, M. R., Liu, T., & Wolfe, J. M. (2012). Reconsidering Yarbus: A failure to predict observers' task from eye movement patterns. *Vision research*, 62, 1-8.
- He, J., Chaparro, A., Nguyen, B., Burge, R. J., Crandall, J., Chaparro, B., ... & Cao, S. (2014). Texting while driving: Is speech-based text entry less risky than handheld text entry?. *Accident Analysis & Prevention*, 72, 287-295.
- Hickman, J. S., Hanowski, R. J., & Bocanegra, J. (2010). Distraction in commercial trucks and buses: Assessing prevalence and risk in conjunction with crashes and near-crashes.
- Hone, K. S., & Graham, R. (2000). Towards a tool for the subjective assessment of speech system interfaces (SASSI). *Natural Language Engineering*, 6(3&4), 287-303.
- Just, M. A., & Carpenter, P. A. (1976). Eye fixations and cognitive processes. *Cognitive psychology*, 8(4), 441-480.
- Klauer, S. G., Dingus, T. A., Neale, V. L., Sudweeks, J. D., & Ramsey, D. J. (2006). The impact of driver inattention on near-crash/crash risk: An analysis using the 100-car naturalistic driving study data (No. HS-810 594).
- Lee, J. D., Caven, B., Haake, S., & Brown, T. L. (2001). Speech-based interaction with in-vehicle computers: The effect of speech-based e-mail on drivers' attention to the roadway. *Human Factors*, 43(4), 631-640. doi:10.1518/001872001775870340
- Lo, V. E. W., & Green, P. A. (2013). Development and evaluation of automotive speech interfaces: useful information from the human factors and the related literature. *International Journal of Vehicular Technology*, 2013.
- McEvoy, S. P., Stevenson, M. R., McCartt, A. T., Woodward, M., Haworth, C., Palamara, P., & Cercarelli, R. (2005). Role of mobile phones in motor vehicle crashes resulting in hospital attendance: a case-crossover study. *bmj*, 331(7514), 428.
- McGehee, D. V. (2014). Visual and cognitive distraction metrics in the age of the smart phone: A basic review. *Annals of advances in automotive medicine*, 58, 15.
- National Highway Traffic Safety Administration. (2012). Visual-manual NHTSA driver distraction guidelines for in-vehicle electronic devices. Washington, DC: National Highway Traffic Safety Administration (NHTSA), Department of Transportation (DOT).

- Parnell, K. J., Stanton, N. A., & Plant, K. L. (2016). Exploring the mechanisms of distraction from in-vehicle technology: The development of the PARRC model. *Safety Science*, 87, 25-37.
- Pernice K. Nielsen J. How to Conduct Eyetracking Studies 2009 Nielsen Norman Group Available at:  
<http://www.useit.com/eyetracking/methodology>
- Pickrell, T. M. (2015). Driver electronic device use in 2013 (No. DOT HS 812 114).
- Ranney, T. A., Baldwin, G. H. S., Smith, L. A., Mazzae, E. N., & Pierce, R. S. (2014, November). Detection response task evaluation for driver distraction measurement application. (Report No. DOT HS 812 077). Washington, DC: National Highway Traffic Safety Administration.
- Road vehicles - Transport information and control systems - Detection-Response Task (DRT) for assessing attentional effects of cognitive load in driving. ISO/DIS 17488
- Sauro, J., & Lewis, J. R. (2011, May). When designing usability questionnaires, does it hurt to be positive?. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 2215-2224). ACM.
- Strayer, D. L., Turrill, J., Coleman, J. R., Ortiz, E. V., & Cooper, J. M. (2014). Measuring cognitive distraction in the automobile II: Assessing in-vehicle voice-based interactive technologies.
- Strayer, D. L., Cooper, J. M., Turrill, J. M., Coleman, J. R., & Hopman, R. J. (2015) The smartphone and the driver's cognitive workload: A comparison of Apple, Google, and Microsoft's intelligent personal assistants. AAA Foundation for Traffic Safety.
- Stutts, J., Reinfurt, D., Staplin, L., & Rodgman, E. (2012). The role of driver distraction in traffic crashes. 2001. AAA Foundation for Traffic Safety.
- Victor, T., Dozza, M., Bärghman, J., Boda, C. N., Engström, J., & Markkula, G. (2014). Analysis of Naturalistic Driving Study Data: Safer Glances, Driver Inattention, and Crash Risk.
- Wickens, C. D. (2002). Multiple resources and performance prediction. *Theoretical issues in ergonomics science*, 3(2), 159-177.
- Wilson, F. A., & Stimpson, J. P. (2010). Trends in Fatalities From Distracted Driving in the United States, 1999 to 2008. *American Journal of Public Health*, 100(11), 2213–2219.  
<http://doi.org/10.2105/AJPH.2009.187179>

- Yarbus, A. L. (1967). Eye movements during perception of complex objects (pp. 171-211). Springer US.
- Young, M. S., & Stanton, N. A. (2002). Malleable attentional resources theory: a new explanation for the effects of mental underload on performance. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 44(3), 365-375.
- Young, K. L., & Lenné, M. G. (2010). Driver engagement in distracting activities and the strategies used to minimise risk. *Safety Science*, 48(3), 326-332.

## Appendix I: The Driving Activity Load Index (DALI)

Question (1 = I do not agree, 7 = I agree)	1	2	3	4	5	6	7
1. The task required my attention	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
2. The task required visual demand	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
3. The task required auditory demand	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
4. The task required tactile demand	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
5. The task required temporal demand	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
6. It was hard to focus on driving while interacting with the system	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
7. I felt stressed using the system	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

## Appendix II: The System Usability Scale (SUS)

Question (1 = I do not agree, 7 = I agree)	1	2	3	4	5	6	7
1. The interaction with the system is consistent	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
2. It is clear how to interact with the system	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
3. It is easy to learn to use the system	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
4. I would use this system	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
5. I felt in control of the interaction with the system	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
6. I felt confident using the system	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
7. The system is easy to use	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
8. I always knew how to use the system	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
9. The system is simple	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
10. I found the various functions in the system were well integrated	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

## Appendix III: The Subjective Assessment of Speech Interfaces

Question (1 = I do not agree, 7 = I agree)	1	2	3	4	5	6	7
1. The system makes few errors	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
2. I was able to recover easily from errors	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
3. I felt tense using the system	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
4. I felt calm using the system	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
5. A high level of concentration is required when using the system	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
6. The interaction with the system is frustrating	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
7. I sometimes wondered if I was using the right word	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
8. It is easy to lose track of where you are in an interaction with the system	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
9. The system responds too slowly	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>