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

Sofia Lindvall

Tutor, Annika Silvervarg
Examiner, Arne Jönsson
Copyright

The publishers will keep this document online on the Internet – or its possible replacement – for a period of 25 years starting from the date of publication barring exceptional circumstances.

The online availability of the document implies permanent permission for anyone to read, to download, or to print out single copies for his/hers own use and to use it unchanged for non-commercial research and educational purpose. Subsequent transfers of copyright cannot revoke this permission. All other uses of the document are conditional upon the consent of the copyright owner. The publisher has taken technical and administrative measures to assure authenticity, security and accessibility.

According to intellectual property law the author has the right to be mentioned when his/her work is accessed as described above and to be protected against infringement.

For additional information about the Linköping University Electronic Press and its procedures for publication and for assurance of document integrity, please refer to its www home page: http://www.ep.liu.se/.

© Sofia Lindvall
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 understood. 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 perceived 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 difference 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

1 Introduction ........................................... 1  
   1.1 Aim and research questions ...................... 2  
   1.2 Limitations and delimitations .................. 3  

2 Background ............................................ 4  
   2.1 Cognitive load and attention .................... 4  
   2.2 Driver distractions and accidents ................ 5  
   2.3 Speech interaction in vehicles .................. 8  

3 Methods ............................................... 10  
   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  

4 Results ................................................ 20  
   4.1 Cognitive load .................................... 20  
   4.2 Visual distraction ................................ 27  
   4.3 User acceptance and efficiency .................. 36  
   4.4 Summary .......................................... 40  

5 Discussion ............................................ 42  
   5.1 Results .......................................... 42  
   5.2 Method ........................................... 44  

6 Conclusions and further research ................... 48
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 its 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 due to time limitations.
2 Background

This section contains a description of theories regarding 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 the following four elements:

- the visuospatial sketchpad
- the phonological loop
- the central executive
- 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 (Sternerg, 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. Further, 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 do not 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 lead to 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 highlight that driver distraction occurs when the drivers are shifting attention away from their primary task - driving. The AAAFTS and Lee et al. (2008) include the concept of safety in their definitions meaning that distraction occurs when the inattention is critical for safe driving or affects recognition of information needed to be 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 come from tasks 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 will 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 be 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 emphasis 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 follow-
ing 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 some speech interfaces could be de-
signed in a way requiring visual-manual demand, even though the main in-
teraction is with speech. Reimer and Mehler (2013) conducted an on-road
study where they could confirm this. Their findings show that implementa-
tions 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 driv-
ing 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 com-
parison to just driving and using visual-manual interfaces.
3 Methods

This section describes the methods 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 systems 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 were 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 Hällereds 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.
4. Tell us the next time you need to take a break.
5. Check your warning messages, vehicle message 2.

For the reminder part of 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). The tasks were chosen based on the results from a user study conducted within the project. The tasks that were chosen were covering identified needs of the drivers.

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 practice 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 were 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 were 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 balanced so that every other is positive and negative.

The items are divided into six main factors: System Response Accuracy, Likability, 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. Likability 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 comple-
ment 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 the audio was 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 of the tasks 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 Hällereds 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 demographic 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 equipment was turned on. The participant begun 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. 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 were a short interview was conducted.

3.8 Data analysis

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 response 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 examination 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-Wilk 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 (because the scale used was a 1-7 Likert scale) 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, Likability, 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 (RT) and hit rate (HR) for the baseline (B) condition and the two interfaces, speech (S) and visual-manual (VM), for each participant.

<table>
<thead>
<tr>
<th>Participant</th>
<th>B (RT)</th>
<th>S (RT)</th>
<th>VM (RT)</th>
<th>B (HR)</th>
<th>S (HR)</th>
<th>VM (HR)</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>506.42</td>
<td>791.6</td>
<td>778.82</td>
<td>1</td>
<td>0.52</td>
<td>0.38</td>
</tr>
<tr>
<td>5</td>
<td>361.1</td>
<td>466.85</td>
<td>682.05</td>
<td>0.97</td>
<td>0.94</td>
<td>0.87</td>
</tr>
<tr>
<td>6</td>
<td>337</td>
<td>564.49</td>
<td>846.69</td>
<td>1</td>
<td>0.85</td>
<td>0.69</td>
</tr>
<tr>
<td>7</td>
<td>269.61</td>
<td>465.02</td>
<td>684.34</td>
<td>1</td>
<td>0.93</td>
<td>0.77</td>
</tr>
<tr>
<td>8</td>
<td>407.38</td>
<td>898.18</td>
<td>995.95</td>
<td>1</td>
<td>0.88</td>
<td>0.88</td>
</tr>
<tr>
<td>9</td>
<td>362.39</td>
<td>441.54</td>
<td>601.76</td>
<td>1</td>
<td>0.97</td>
<td>0.77</td>
</tr>
<tr>
<td>10</td>
<td>624.59</td>
<td>634.61</td>
<td>943.16</td>
<td>0.97</td>
<td>0.91</td>
<td>0.88</td>
</tr>
<tr>
<td>11</td>
<td>376.89</td>
<td>709.24</td>
<td>1001.43</td>
<td>0.97</td>
<td>0.74</td>
<td>0.50</td>
</tr>
<tr>
<td>12</td>
<td>326.5</td>
<td>491.2</td>
<td>713.39</td>
<td>0.98</td>
<td>0.88</td>
<td>0.85</td>
</tr>
<tr>
<td>13</td>
<td>478.73</td>
<td>747.55</td>
<td>960.15</td>
<td>1</td>
<td>0.8</td>
<td>0.56</td>
</tr>
<tr>
<td>14</td>
<td>625.59</td>
<td>987.34</td>
<td>997.76</td>
<td>0.85</td>
<td>0.71</td>
<td>0.62</td>
</tr>
<tr>
<td>TOTAL</td>
<td>425.11</td>
<td>654.60</td>
<td>836.86</td>
<td>0.98</td>
<td>0.83</td>
<td>0.71</td>
</tr>
</tbody>
</table>
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, $F(2, 20) = 59.89$, $p < .001$, $\eta^2_p = .86$. Mauchly’s test indicated that the assumption of Sphericity was met, $x^2(2) = .69$, $p > .05$. 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 1: The response time for the three different conditions.
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, $F(2, 20) = 19.61, p < .001, \eta^2_p = .66$. 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, $F(2, 20) = 18.27, p < .001, \eta^2_p = .65$. 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-
significant 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**

![Graph showing hit rate](image)

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, \( x^2(2) = 21.54, p < .001 \). 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 (M\( \mu \)d = 1.00, IQR = 0.03) had a statistically significantly higher hit rate than both speech (M\( \mu \)d = 0.88, IQR = 0.19) (\( z = -2.936 \), \( p = .003 \), \( r = -0.55 \)) and visual-manual interaction (M\( \mu \)d = 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, \( \chi^2(2) = 11.023 \), \( p = .004 \). A post hoc test using Wilcoxon signed ranks test revealed that baseline (M\( \mu \)d = 1.00, IQR = 0.03) had a statistically significantly higher hit rate than the speech (M\( \mu \)d = 0.88, IQR = 0.19) (\( z = -2.93 \), \( p = .003 \), \( r = -0.55 \)) and visual-manual interface (M\( \mu \)d = 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, \( \chi^2(2) = 16.79 \), \( p < .001 \). A post hoc test using Wilcoxon signed ranks test revealed
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, \( x^2(2) = 15.85, p < .001 \). 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).
The Friedman test revealed that there was a statistically significant difference in raw DALI-score for the tasks, $x^2(2) = 17.72$, $p < .001$. 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

![Graph showing fixation on road for three conditions: Baseline, Speech, Visual-Manual. Error bars are included for 95% CI, and the median (Mdn) and interquartile range (IQR) are provided.]

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, $x^2(2) = 16.22, p < .001$. 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, $x^2(2) = 14.89, p = .001$. 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, 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, $x^2(2) = 6.22$, $p = .045$. The post hoc test revealed that the speech interface ($M_{nd} = 16.00$, IQR = 17.50) had a statistically significantly lower amount of glances on task displays compared to the visual-manual counterpart ($M_{nd} = 32.00$, IQR = 18.00) ($z = -2.43$, $p = .015$, $r = -0.61$). The difference between the baseline ($M_{nd} = 23.00$, 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, $x^2(2) = 16.22$, $p < .001$. A post hoc test with the Wilcoxon signed rank test revealed that baseline ($M_{nd} = 23.00$, IQR = 18.00) had a significantly higher amount of glances on task displays compared to the speech interface ($M_{nd} = 8.00$, 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, $x^2(2) = 14.82$, $p = .001$. 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 (Mnd = 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), $(z = -2.67, p = .008, r = -.57)$. 33
Per task

![Figure 11: Glances on task displays with a duration of 2 sec or more for the three different conditions per task.](image)

**Task 1.** The Friedman test revealed an overall significant effect of condition on number of glances with a duration of 2 seconds or more, $x^2(2) = 14.00$, $p = .001$. A post hoc test with the Wilcoxon signed rank test revealed that there was a significant difference between the baseline (Mdn = 0, IQR = 0.50) and visual-manual interface (Mdn = 4.00, IQR = 6.50) ($z = -2.54$, $p = .011$, $r = -0.64$) as well as between the speech (Mdn = 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, $x^2(2) = 13.94$, $p = .001$. 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 (Mdn = 7.00, IQR = 4.50), ($z = -2.67$, $p = .008$, $r = -0.67$). A significant effect was also found between the speech (Mdn = 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 $x^{2}(2) = 15.25$, $p < .001$. 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), \( (z = -2.85, p = .004, r = -0.61) \).
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 likability. *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 (Mnd = 6.00, IOR = 3.00) than for the visual-manual interface (Mnd = 3.00, IQR = 2.00), (z = -2.66, p = .008, r = -0.57). *Habitability* had a significantly
higher value for the speech interface ($M_{nd} = 4.67$, $IQR = 1.00$) than for the visual-manual interface ($M_{nd} = 2.67$, $IQR = 1.33$), ($z = -2.40$, $p = .016$, $r = -0.52$). The factor likability was statistically significantly higher for the speech interface ($M_{nd} = 5.80$, $IQR = 1.20$) than for the visual-manual interface ($M_{nd} = 4.00$, $IQR = 1.40$), ($z = -2.85$, $p = .004$, $r = -0.61$).

**Interviews**

This sections presents the themes and results from the interviews.

**System Response Accuracy**

Four of the participants 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 systems, 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

Measurements 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 of 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 critically discusses and elaborates on the answer that has been given to each research question and then turns to a discussion of the adequacy of their assessment.

5.1 Results

How does the speech interface as compared to the visual manual interface affect distraction while driving?

The results for distraction deriving from cognitive load will be discussed first. The first analysis looked into the 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 when driving than compared to the visual-manual interface. Furthermore, the results from the interviews show that several of the truck drivers explicitly expressed that they felt that speech system let them have their attention on the roadway while they were driving.

The next analysis looked at each task separately. The objective measurement of cognitive load indicated that the navigation task was less demanding when using the speech interface when compared to the visual manual interface. This was also the task were the results for the speech interface was closest to the results for the baseline tasks with regard to hit rate data. However, the calling task, playing a song, and noting a reminder did not reveal any significant differences in the comparison between the interfaces. When looking into mean response times, the visual-manual interface measured 825.61, 917.05 and 918.94 ms for the three tasks respectively. In comparison, the speech interface measured 691.98, 749.49 and 596.16 ms. The results showed that the response times in the different task conditions were quite similar in the visual-manual interface evaluation. In the speech interface evaluation, response times were significantly higher for task 1 and 2 than for task 3, which is the navigation task. Lee et al. (2001) showed that more complex speech tasks impose a higher demand on 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 seemed to be less complex
when done through the speech interface.

The assessment of visual distraction show that the speech interface is less visually distracting than the visual-manual interface for all measurements and tasks (except in number of glances for task 1). The speech interface let the drivers have a higher fixation on the roadway and demanded fewer glances on tasks displays. Important to note is that these results were found despite the fact that the speech system gave visual stimuli to the driver. As mentioned before, the calling task was the only task where the visual-manual counterpart did not require significantly more glances than the speech system. A reason for this could be that the drivers are so used to carrying out the task that it has become almost automated.

Is interaction via voice a safer and more efficient way to communicate compared to traditional interfaces?

Several researchers have found a strong connection between unsafe driving behaviour and visual distraction (for example Hickman et al., 2010; Engström, 2011; Victor et al., 2014) which indicates 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. According to the NHTSA (2012), the measurement with the strongest connection to safety is to register off-road glances greater than 2 seconds. These glances should be avoided since they have been proven to increase crash risk. For this measurement, the visual-manual interface had a significantly higher number of glances for all tasks, which further suggests 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 to a lesser extent 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 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 in comparison to the visual-manual system. For SUS, the median score is 75.02 which is above 68 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 68 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 likability 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 Hällereds 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 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, thus 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. It is however 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. Interested 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. However, because of time limitations the measurements could not be added in this project.
6 Conclusions and further research

A comparison has been conducted between a speech interface and its visual-manual counterpart with regard to driver distraction, safety, and user acceptance and perceived efficiency. 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 perceived 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


49

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.


Road vehicles - Transport information and control systems - Detection-Response Task (DRT) for assessing attentional effects of cognitive load in driving. ISO/DIS 17488


Wickens, C. D. (2002). Multiple resources and performance prediction. Theoretical issues in ergonomics science, 3(2), 159-177.


51
Yarbus, A. L. (1967). Eye movements during perception of complex objects (pp. 171-211). Springer US.
### Appendix I: The Driving Activity Load Index (DALI)

<table>
<thead>
<tr>
<th>Question (1 = I do not agree, 7 = I agree)</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. The task required my attention</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. The task required visual demand</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. The task required auditory demand</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. The task required tactile demand</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. The task required temporal demand</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. It was hard to focus on driving while interacting with the system</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. I felt stressed using the system</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Appendix II: The System Usability Scale (SUS)

<table>
<thead>
<tr>
<th>Question (1 = I do not agree, 7 = I agree)</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. The interaction with the system is consistent</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. It is clear how to interact with the system</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. It is easy to learn to use the system</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. I would use this system</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. I felt in control of the interaction with the system</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. I felt confident using the system</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. The system is easy to use</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. I always knew how to use the system</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9. The system is simple</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10. I found the various functions in the system were well integrated</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Appendix III: The Subjective Assessment of Speech Interfaces

<table>
<thead>
<tr>
<th>Question (1 = I do not agree, 7 = I agree)</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. The system makes few errors</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. I was able to recover easily from errors</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. I felt tense using the system</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. I felt calm using the system</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. A high level of concentration is required when using the system</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. The interaction with the system is frustrating</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. I sometimes wondered if I was using the right word</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. It is easy to lose track of where you are in an interaction with the system</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9. The system responds too slowly</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>