CHALMERS

Nuisance vs False Collision Warnings During Driving

The Difference in Effect of False and Nuisance Alarms on the Driver when Driving with a Forward Collision Warning System

Master of Science Thesis in Interaction Design

ERIK H. HALLGREN

Department of Computer Science and Engineering
CHALMERS UNIVERSITY OF TECHNOLOGY
UNIVERSITY OF GOTHENBURG
Göteborg, Sweden, June 2009

Nuisance vs False Collision Warnings During Driving
The Difference in Effect of False and Nuisance Alarms on the Driver when Driving with a Forward Collision Warning System

ERIK H. HALLGREN

Examiner: FANG CHEN

Department of Computer Science and Engineering
Chalmers University of Technology
SE-412 96 Göteborg
Sweden
Telephone + 46 (0)31-772 1000

Department of Computer Science and Engineering
Göteborg, Sweden, 2009

Abstract

As a diverse field of advanced driver assistance systems (ADAS) are developed and incorporated into the driving environment the need for analysis of their effects on the driver becomes increasingly important. As ADAS gain more influence on the driving task they change how we drive with a great potential for increasing the safety of the driver and those in his/her vicinity. There is however also a potential problem with the increasing amount of aids, if tuned incorrectly they may decrease safety instead; it is therefore of vital importance that the effects of aids and how they relate to potentially hazardous situations is well understood.

This thesis strives to provide a contribution to the research field of ADAS in order to increase safety in the driving environment, specifically in the area of trust in frontal collision warning (FCW) systems under different circumstances. Tests were carried out in a solid base driving simulator where reaction times were measured, followed by subjective measurement of workload by means of NASA-TLX questionnaire and subjective ratings of trust by means of a novel questionnaire based on the APT-framework[12].

The test was between groups by design where one group drove a scenario using a FCW system prone to giving nuisance alarms and the other group a similar scenario with a FCW system prone to false positive alarms. The FCW used was similar to that used in Volvo's cars. Both scenarios presented the test subject with 6 alarm situations, of which two where different between the scenarios. Each scenario proceeded for 12-14 minutes depending on speed and reactions to events.

No significant differences was found between the two groups in any area of data extracted. It is concluded that there is no difference in impact of nuisance and false positive alarms in either of the areas of performance, workload or trust as they were constructed here. There is evidence of the participants rating the type of alarms differently between the groups, supporting this conclusion. Reaction times lie well within what can be considered normal during real driving, indicating that the data found can be useful in further studies in the area of ADAS.

The results of this study needs to be further tested but it carries some implications for the algorithms used in FCW systems in what types of alarms should be triggered. A proposed continuation of this study is in the relation between true positive alarms and nuisance alarms to further define the role of nuisance alarms, as well as a longitudinal study showing the long term effects of nuisance alarms.

KEYWORDS: Trust, Cooperation, FCW, ADAS, False Alarms, Nuisance Alarms, Simulator, NASA-TLX, Secondary Task

Preface

This master thesis constitutes the concluding work of the master program in Interaction Design at Chalmers University of Technology in Göteborg. The thesis work has been conducted at SAFER with Volvo Cars AB as commission assigner. The intended readers are researchers dealing with active safety systems, people at Human Machine Interface departments and others that are curious about the focus of this master thesis.

Special thanks to:
Henrik Lind
Fang Chen
Anders Lindgren
Peter Ljungstrand
Alexander Angelelli
Paul Alvarado Mendoza
Oskar Wenneling
Daniel Engström
Niklas Andersson
Claes Andersson
John D. Lee
Jacques Del Mar

Contents

List of Figures

List of Tables

1 Introduction

The modern car is a vehicle full of intricate systems of a diverse nature controlled by a computer that is mostly controlled by the driver. In traditional mechanical car control on the other hand, human action corresponded to reaction in a direct way without interceding interpretation, i.e. pressing the brake pedal caused the brake discs to press against the wheel through hydraulics. The modern car enables systems far more advanced than those possible in the mechanical version, these help the driver take correct actions as well as force the driver to accept giving up some degree of control.

As more and more automation and decision aids are implemented in the car, novel problems dealing with how a driver will respond to and accept the various systems appear, to which we will soon return.

Automation and decision aids can be described as part of a sliding two-dimensional scale with full automation at one end and lack of automation at the other. Automation is here regarded as automation of control over the car that the human driver normally carries out, not clutch, gas, servos and so on that translates the human action into commands in the car's computers.

In this interpretation full automation equals an autonomous car without need for human interaction when operating, able to safely get from point A to B completely by itself. On the other hand a human driver is not able to safely go from point A to B without support from the car, the driver needs to know what speed he or she is travelling since at higher speeds this is impossible to comprehend with sufficient accuracy, he or she needs to know whether there is enough gas for the trip and so on. All this information has to be provided by the designers of the car and the human driver has to rely on this information in order to use the car effectively, the different dials and lights in the car can therefore be seen as simple decision aids, enabling an informed decision by the driver. Even though these aids show simple measurements they may be inexact like e.g. many speedometers, however the driver adapt to them and learn to estimate the correct value. The fuel indicator doesn't show an appropriate amount of fuel unless you are driving at an even speed on a flat road, this is usually not a big problem unless you constantly drive in slopes or like to drive without margins. What these examples show are that even the simplest decision aid has problems with different implications to the driver; such as forcing an extra effort of interpretation of the aids.

Most modern cars have some degree of automation installed. Enabling the driver to focus on other things than e.g. keeping the speed on a motorway has both advantages and disadvantages. The advantages of giving up some of the control to a computer involves relieving stress and workload, enabling the driver to focus cognitive resources on other parts of the driving or driving related activities such as using the In-Vehicle Information System (IVIS) to keep track of a GPS-system for navigation [54]. This example enables security-increasing behavior such as being able to scan a dark road for wild animals while listening to directional voice cues instead of looking at a paper map. Disadvantages of automation are that the driver may focus on non-driving related tasks such as changing radio stations or talking to other passengers as well as deteriorate in driving skills as the system takes over more tasks [53]. As driving is a cognitively intense task, the cognitive resources of the driver used for driving are reduced, causing an increased risk of errors. In order to cope with a higher cognitive load the driver divides his or her attention by observing the environment with shorter glances instead of scanning it continuously for possible problems [26].

As we shall see in this thesis, these problems can be quite complex and furthermore very crucial to the safety of the driver as well as the performance of the system, two parts inherently relevant to each other, especially when it comes to safety related systems.

2 Purpose of Study

The goal of the study carried out in this thesis is to examine an Frontal Collision Warning (FCW) interface with the purpose to discern how different degrees of correctness in the warnings affects the driver in the aspects of trust to the system. Furthermore it will be studied if the trust has a correlation to the effectiveness of the FCW in the form of changed response to the warning signal. This is instrumental in the acceptance of the system by the user and more importantly, a high percentage of incorrect warnings can cause more problems to the driver than driving without the warning system and may therefore become a safety hazard.

Knowledge of the effect on reaction time from FCW exposure, both with and without false alarms, has implications in driver modeling and when designing a warning system and could lead to a better model of the driver reaction. This in turn could lead to a more efficient warning algorithm and thus a different tuning of the system, which would bring safety benefits as well as carry research on the topic forward.

The FCW interface in the study is similar to the FCW interface in Volvo vehicles consisting of a multimodal interface, combining visual and auditory cues.

2.1 Connection to previous studies

In an article from 2007 [31] a comparison was carried out between four different types of visual Forward Collision Warning (FCW) displays with the goal of discerning which would yield the lowest reaction time by the driver. A heads-up display (HUD) was found to be significantly more effective in this regard than the other types tried (in the steering wheel, on top of instrument panel with icon and positioned within the instrument cluster with icon). This thesis continues the research by further examining the effects of this type of warning system on the driver, leading to insights in the impact of calibration of the algorithm behind the warning system.

2.2 Hypotheses and research questions

- *Null hypothesis: There will be no difference in reaction times between the two test groups as well as no difference in workload or trust in the system*
- *Alternate hypothesis: Drivers driving with a nuisance*[1] *prone FCW will have significantly better reaction time as well as having lower workload and a higher degree of trust*

[1] *The definition of nuisance alarm is in this thesis defined as a correct alarm that is understood by the driver though the following situation did not lead to a critical event. A false alarm is not understood by the driver. Typically a nuisance alarm is a car braking to turn while a false alarm is an alarm that is incorrectly triggered by guard rail outside of the road without apparent reason.*

Figure 1: Volvo's FCW HUD

This thesis will also strive to answer the research questions posed below:

- What are the effects of false alarms versus nuisance alarms in a driving environment?
- How will the driver's subjective estimate of trust differ between the different scenarios?

2.3 Delimitations

A stationary based simulator was used to be able to accurately repeat scenarios and enabled measuring test person responses in a controlled environment. The scenario was set up using moderately curved roads with low lateral acceleration to reduce simulator sickness.

An interview was planned but cancelled due to time constraints. The interview would have been carried out with existing users and focus on their understanding of and trust in the FCW to be compared with results from the simulator tests; this would have provided a longitudinal perspective which would have been desirable though it would have been a comparison between quantitative and qualitative data. The design of the simulator study was altered with regards to the questionnaire due to the lack of need to do this comparison.

Sex, age and driving experience was not controlled in the choice of participants; the study was limited to 29 test participants due to time constraints in the simulator.

Six situations of a limited selection were presented to the driver and in a country road scenario with a speed limit of 90 km/h, taking approximately 12 minutes. Under these conditions during real driving it would take a minimum of 40 hours to experience the same amount of warnings.

3 Literature review

An introduction to the technology, the human factors research involved with the technology and background to the prevalent subject of this thesis.

3.1 Advanced Driver Assistance Systems

Advanced Driver Assistance Systems (ADAS or ADA Systems) is a collective term denoting the various active safety systems implemented in the driver environment in order to decrease the risk of accident. It would be meaningless to list all the different types as they are constantly being developed, examples commonly provided by most car manufacturers around the world include Adaptive Cruise Control (ACC), Collision and Lane Departure Warning Systems and Blind Spot Information Systems[32]. ADAS is not to be confused with the neighboring and sometimes overlapping set of technology called In-Vehicle Information Systems (IVIS) which concern different types of information presented in the driving environment but not necessarily safety related as such, for example GPS or text messaging.

The benefits of correctly implemented and designed ADA Systems can have a large impact upon society at several levels. According to Eurostat, 47,000 lives were lost in road accidents, 1.7 million were injured, note that not all accidents reported include a car, however, the vast majority does [1], reducing these numbers would reduce human suffering as well as costs involved with the treatment of traffic victims, around 70 billion Euro's [10]. Other potential gains includes reduction of traffic congestion and more fuel efficient driving which would reduce pollution as well as result in both health related and economic benefits [13].

Several potential safety benefits have been shown to come from the implementation of ADAS. The implementation of Electronic Stability Control, which detects and attempts to prevent the vehicle from skidding, in the U.S has been shown to potentially being able to save 7000 lives per year if all passenger vehicles were equipped with it [21]. Other technologies have more unclear benefits such as ACC, where there are potential benefits such as reduction of tailgating accidents and more fuel efficient driving [17], as well as drawbacks in the form of complacency to the system which may result in reduced vigilance[54].

Often implemented together with the ACC is the Forward Collision Warning (FCW) system as this uses the same radar technology as ACC. FCW is increasingly being implemented in newer vehicle models from car manufacturers. When the driver encounters a potential imminent crash situation he/she is warned and ideally has an increased chance of avoiding the situation.

Three main areas of interest have been researched with respect to the FCW: cue modality, cue presentation and accuracy and will be further detailed here.

- Cue modality

Liu [33]tested the difference between a multi-modal cue, an auditory cue and a visual cue in the context of information intense displays and found that the multi-modal cue was superior, followed by the auditory cue and last the visual, it was not clear however whether the inferior performance of the visual cue may be caused by a bad choice of presentation or due to its modality. The multi-modal cue consisted of a combination of the visual and auditory cues in conjunction. Tijerina et al tested three different types of haptic cues in a rear end collision warning system [45] similar in function to the forward collision warning used in this thesis. They compared the effects of an active steering display, i.e. vibration of the steering wheel, to a brake pulse display, i.e. the warning system signaling through taking over the break system in a short pulse. Concluding that a breaking mono-pulse may be beneficial, they are reserved but positive in their recommendation of such a display. Lee et Hoffman states that the advantage of a mono-pulse breaking display vs. a vibration type display comes from the natural mapping of breaking vis-à-vis the collision situation which the vibration does not have [28]. There was no evidence of better performance with an auditory than haptic modality though the haptic display was preferred subjectively.

- Cue presentation

There are two ways in which the cue can be presented, a Likelihood Alarm Display (LAD) where the degree of danger is shown [43], or a display with two possible states, on or off. The study by Lee et al, referenced above, found support for a LAD type display over a two state display in the areas of trust, performance and lack of incorrect response in collision warning systems. There is however a danger in using LAD that the cognitive load may be increased to the extent that it outweighs the positive effects.

- Accuracy

Algorithms controlling when the system will be triggered needs to be accurate as the ratio of false alarms vs. correct alarms affects the driver in several ways, this is discussed further below.

3.2 Trust and Corporation

Trust deeply affects the way we choose to interact with the world and those around us, it can be seen as the basis for all our actions; from walking (trust in our muscular control vs. the structure of the ground) to making life-altering decisions (trust in our judgment as well as the circumstances affecting it). We may have doubts in what the result will be and thus have to weigh the cost of being wrong against the advantage gained by being right. Affecting trust however is, for example, wishful thinking that also forms the basis of our actions even though it is ungrounded, as opposed to knowledge of the relevant situation. Trust, as discussed here, is most relevant due to its connection to the act of cooperation, as cooperation between agents is at the heart of deciding whether to conform to or oppose a proposition. In order to reach a state of cooperation a threshold of trust need to be reached between the involved agents, this threshold is unique to every situation and may be difficult to discern [19]. Reeves & Nass[40] shows that the concept of trust between two humans carry over to trust between human and machine to a very high degree allowing conclusions in this area to be drawn from social psychology, where this subject have been studied extensively. Lee et See's definition states that

> "Trust can be defined as the attitude that an agent will help achieve an individual's goals in a situation characterized by uncertainty and vulnerability. In this definition, an agent can be automation or another person that actively interacts with the environment on behalf of the person" [29].

Innate to cooperation is the agent's conceptual model of the system with which he or she interacts, if this model is based on faulty conclusions or data the cooperation between system and agent will falter as the agent will not be able to assess the situations correctly [36]. Successful cooperation depends to a great extent on how the system acts against the environment where it is working, the agent's conceptual model is a delicate thing and is easily affected by false alarms and misses by the system. Both over trust and mistrust on the agent's behalf is probable in situations where he or she is uncertain of how the system works, i.e. the agent's estimate of how correct the system is differs from the real value [38].

3.3 The effect of imperfect aids

Collision warning alarms have been the subject of a number of studies, many in aviation settings which are where they were originally developed but more recently studies in car settings have been carried out to a greater extent. Most of the studies which were carried out with focus on collision warnings in cars focus on the nature of the warning signal, trying to heighten the sense and immediacy of the alert presented to the driver in order to cause the right reaction. Results from aviation studies and other fields on the subject of effects from false warnings and such can be generalized to the automotive environment as the nature of the warning is safety critical in both instances and the same cognitive abilities and agent behavior are involved. There does not seem to be a greater difference in results between types of tasks than between experiment settings.

Testing different reliabilities in a collision warning aid, Bliss et Acton showed that the level of trust decreased with the reliability of the aid by measuring collision rate, appropriateness of reaction and degree of swerving to avoid collision [8]. Reliability here being exclusively focused on FA and not misses, these findings are supported by Enriquez et al [18].

Studies in aviation settings carried out by Dixon et al indicates that false alarms and misses affects the user in different ways, they used a detection task where the test subjects knew what type of warning problems would occur and also whether the aid was perfect or not perfect [16]. They divided agent interaction with a warning aid in compliance and reliance according to a model of imperfect automation introduced by Meyer, in Meyer's model FA affects compliance and FN affects reliance exclusively [35]. Compliance is the extent to which the agent follows a presented warning and to what degree he or she prioritizes the warning over other tasks, reliance on the other hand is the extent to which the agent trusts that a warning will be presented when appropriate. According to Dixon et al compliance is selectively affected by FA to a high degree while reliance is affected by both FA and FN, these findings are supported by Wickens et al [50]. This leads to the conclusion that depending on how the aid is designed, if it is more prone to either FA or FN, it will affect the behavior of the agent operating with the warning aid in different ways. If the threshold of either FA or FN is too low it will cause a worse performance than with no warning aid at all as the agent either has to analyze the data if he or she is prone to compliancy or will be distracted and annoyed by the alerts if prone to reliance. A threshold of lower than approximately 70% correct warnings has been shown to cause more problems than not having a warning aid at all, this is explained by the fact that the agent even when aware of a system's high degree of flaws uses it to reduce cognitive load in order to focus on other areas of operation [49]. It is important to point out that the percentage is not a definitive figure as it is an estimate based on a collection of studies and does not address the difference between FA and FN prone systems, furthermore, training and experience may compensate for the shortcomings of a system as well as the degree of importance in the aid related task.

The effect of imperfect warning aids becomes increasingly complex when you factor in simultaneous events in situations where the aid is acted upon; Besnard et al. found that an unrelated event sometimes was incorporated into the agent's mental model of how the system works [6]. Implications of this are that the agent needs to be informed in how the warning aid works in order to build a correct mental model of it [24]. Wiese found that warnings from other systems may increase or decrease the performance of the aid with regard to reaction times, an e-mail alert from an IVIS presented 1000ms before a frontal collision warning decreased performance slightly, while when presented 300ms before showed a large decrease [51]. This is due to two concurrent tasks competing for the cognitive abilities of the agent which has the result of decreasing either both or one of the tasks depending on priorities and urgency. The effect can however be reduced when the concurrent alerts are of different modalities which would reduce the bottleneck effect of information processing and the selection of an appropriate response[39].

3.4 Definition of alarm types

When discussing the subject of warning aids there are several ways in which the alarms may be classified: False positive (FP), False negative (FN), positive or negative. Positive and negative alarms are here simply the presentation, or lack thereof, of the alarm when appropriate, FP on the other hand constitutes alarms that are triggered for no reason and FN are situations where the alarm should have triggered but didn't. In the literature on impact of FA on the

Alarm Type	Explanation
True Positive (TP)	Correct alarm presented in a situation which warrants a reaction by the driver to avert collision.
False Positive (FP)	Incorrect alarm presented in a situation which does not warrant a reaction by the driver.
True Negative (TN)	Lack of alarm presentation when not warranted.
False Negative (FN)	Incorrect lack of alarm presentation in a situation that warrants it.
Nuisance Alarm (N)	True positive in the sense that a situation occurred that according to requirement should provide an alarm. However since the situation did not develop into a critical situation the alarm turned out False Positive.

Table 1: Alarm Type definitions

agent, the terms FP and FN are used in their extreme form, as explained above. However as it will be adressed here, this is not a fruitful definition. The relation between a FP and a positive warning and the difference between a FN and a negative warning is not of a discreet nature as there is no possibility of an objective measurement, what is conceived as a FP warning by one operator may be conceived as a positive warning by another due to, for example, differences in operating experience or preconceptions of how a warning system should behave. This is further complicated by the effect that the warning system may have on the driver in itself, where a bias towards or against the warning system affects the operators assessment of the situation [8]. Also affecting the definition of false alarms is the fact that an operator in a complex environment with a high cognitive load will misinterpret some situations, when confronted with a FN he/she may interpret it as a true negative due to lack of allocating resources to the area of emergency [18] and thus give a incorrect response which may cause a critical situation to arise.

It follows that there is a difference between perceived and actual false alarms but that this is hard to discern objectively as the driver's individual skills and preferences defines them. The need therefore arises to define a type of alarm that is defined from the viewpoint of the driver and allows for a broader definition where the alarms are not necessarily neither true nor false, but something in between. Nuisance alarms (N) can be defined as alarms that have a clear connection to the situation where they are presented, and are in this sense true positive, but if the situation does not warrant an alarm, from the viewpoint of the driver or by the situation simply not developing critically, they are at the same time false positive. Xiao et Seagull defines nuisance alarms as: "Nuisance alarms are those indicating state changes that are dangerous to system integrity in some context, but not in the context in which they are set off." [52].

Example of a nuisance alarm situation can be found in a study by Lee et al where the nuisance alarm is defined by a breaking lead vehicle situation but where the vehicle then accelerates again, not warranting the collision alarm presented to the driver [28]. Kiefer et al gives a generalized definition of nuisance alarms and divides them into two classes, In-path and Out-of-path. In-path nuisance alarms are triggered by an object in the driver's headway at too far a distance to warrant an alarm and Out-of-path nuisance alarms are triggered by an object out of the driver's path and therefore don't pose a potential collision threat [25].

3.5 Annoyance

Annoyance, or irritation, is here defined as the subjective experience of interruptions or distractions during task execution which causes discomfort to the driver. No substantial body of research has been found on the subject of interruptions in safety or security critical situations focusing on annoyance and its causes; it is therefore difficult to draw confident conclusions on how the interruptions will affect the users. In the collision warning case we propose that the term Annoyance is related to an excessive amount of false and nuisance alerts that cause discomfort to the driver and could have negative impact on the estimates of performance and may cause the driver to turn the system off.

3.6 Models of the decision process

Viewing trust as a subjective basis for decision making, theories regarding this area can be used in order to model how and why decisions are made and on what foundation. With the exception of the Argument Based Probabilistic Theory below, the following models and theories are here explained with their direct relevance to this thesis and are to be seen as an overview of potential tools for continued studies in the subject of trust and decision making in safety critical situations. The Argument Based Probabilistic Theory warrants a further explanation as it will be applied in this thesis.

3.6.1 Argument Based Probabilistic Theory

The Argument-based probabilistic Theory (APT) [12] provides a model which can be used to explain how and why the driver changes his/her model during the tim of aid use, its predictive power is however limited. It consists of five parameters structured around five components which are inherently linked to each other. Components - grounds, warrant, backing, rebuttals and qualified claim - represents knowledge and belief the user has about the system which can be expressed qualitatively, hence "Argument-based". Parameters - completeness, resolution, reliability, calibration and temporal scope - on the other hand represents different aspects that modifies the components over time, either by

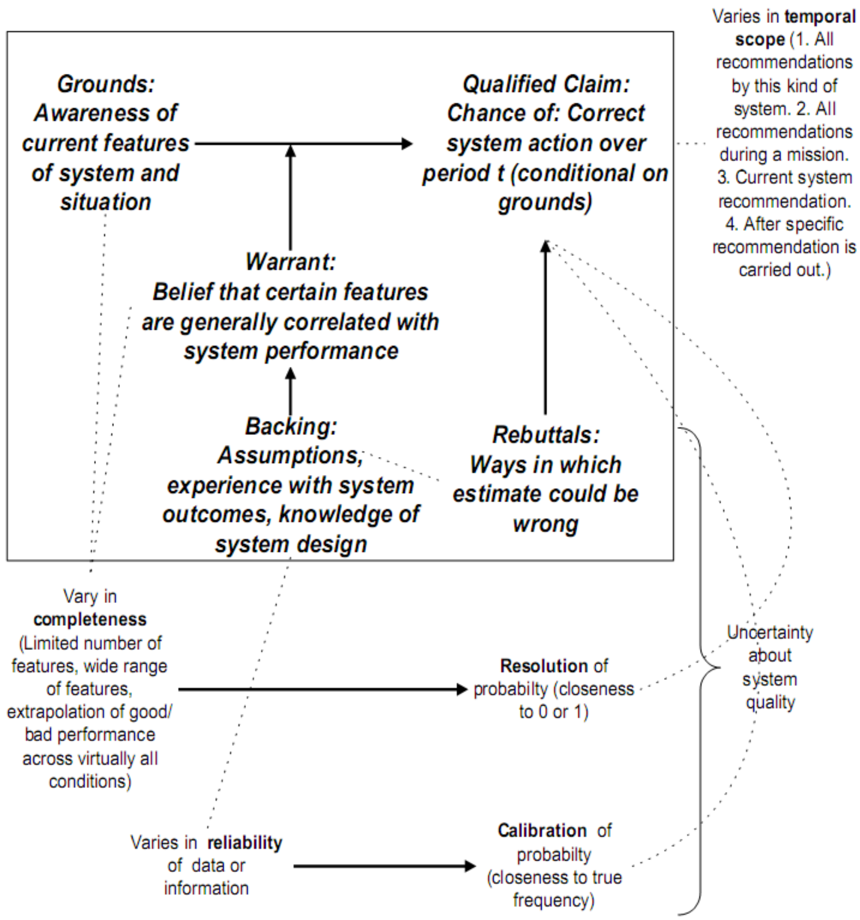

Figure 2: Argument based Probabilistic Theory[12]

empirical experience or new theoretical knowledge of the system in question. The model is shown in figure 2 and detailed below.

Components

The components' part of APT is heavily based on Toulmin's theory of argument which had the goal of examining how arguments in a qualitative context, where Toulmin argues formal logic would be insufficient, are formed [46].

Grounds Grounds represents awareness of the features in the system and the situation in which the system is used, this can be explicit knowledge gained from learning about the system from a source and thus is not reliant on use of the system.

Warrant Rather than awareness of the features, warrant represents beliefs in how the features of the system will react to certain situations. This is most commonly expressed through vague statements such as “highly stable under this condition” but can also be quantitatively stated i.e. “70% correct under these circumstances”.

Backing The origin of the knowledge and beliefs of and in the system constitutes backing. This affects Warrant as different sources of information would yield different strength in opinions about the functioning of the system; e.g. “I know this to be true as I read the manual” vs. “I believe this may be true as that is the opinion I have formed from hearing about these kinds of systems”.

Rebuttals Rebuttals are the believed or known exceptions to the warrant and may be either implicit or explicit. Implicit rebuttals may be part of the backing when baseing a warrant on previous experiences, but may also be explicit assumptions expressed on their own accord. These explicit assumptions can be part of a strategy such as assuming the worst case validity of the warning system until proven otherwise, when contradicted they are however reevaluated and corrected if necessary.

Qualified Claim The components above together lead to a qualified, rational in the sense of being backed up by a process of assessment, claim about whether to reject or accept a warning or even the whole system in a longer perspective.

Parameters

Parameters are used as a way of explaining what affects the components above, as well as how these change over time causing a change in the resulting qualified claim.

Completeness Completeness reflects the degree of understanding the user has of the system under certain conditions at any given temporal phase based on Warrant and Grounds.

Resolution Resolution is the degree to which the user is able to discern the probability of the warning system being right in its recommendation. This relates to the resolution of the warning system in specific situations which enables the user to assess a specific situation based on the completeness of knowledge of the system, i.e. 80% chance of the system being right in one situation but 60% in another.

Reliability Reliability concerns the amount of information and experience that underlies the backing as well as the quality of the information. The more experience the user has with a system , the more reliable his/her trust assessment becomes, someone with little experience may give a system a reliability with a wide scope, i.e. 50%-100% correct while an experienced user with knowledge of the system may give a narrower scope, i.e. 80%-90% correct.

Calibration The difference between expected probability of correctness in the warning system and true probability of correctness is encapsulated by the Calibration parameter. This is related to the situation in Grounds as well as Reliability. Calibration can be expressed as being the chance of being right in ones assessment of a warning in a given situation, for example an experienced user may be off by 5% while an inexperienced user may be off by 25%.

Temporal Scope Temporal scope is divided into four perspectives on time, each with a decreasing temporal scope and with a starting point at the beginning of the defined perspective.

1. Trust in the system over all potential uses.
2. Trust in the system through a specified time period or task.
3. Trust in a specific recommendation or warning by the aid, **before** reacting to or verifying the validity of the recommendation or warning.
4. Trust in a specific recommendation or warning by the aid, **after** reacting to and verifying the validity of the recommendation or warning.

3.6.2 ACT-R

ACT-R, Adaptive control of thought–rational [4], is a complex unified theory of mind which assumes that cognition is divided into modules that are inherently connected to each other. Figure 3 shows a proposed structure or architecture of modules that are central to the ACT-R theory, there is however no defined number of modules that can be implemented as this is ever evolving to emulate new brain functions as needed based on cognitive research, especially in the neuroscience field [3]. The modules are put together and defined in the ACT-R framework which is in the proximity of a programming language where the modules can be seen as programs [11].

The research on ACT-R has been applied to model various fields such as: learning and memory, problem solving and decision making, language and communication, perception and attention and cognitive development. A few examples of applications follows:

- Human-computer interaction to create models of users to automatically assess computer interfaces[44].
- Education (cognitive tutoring systems) to facilitate learning by “guessing” what difficulties student may run into and help them. [30].

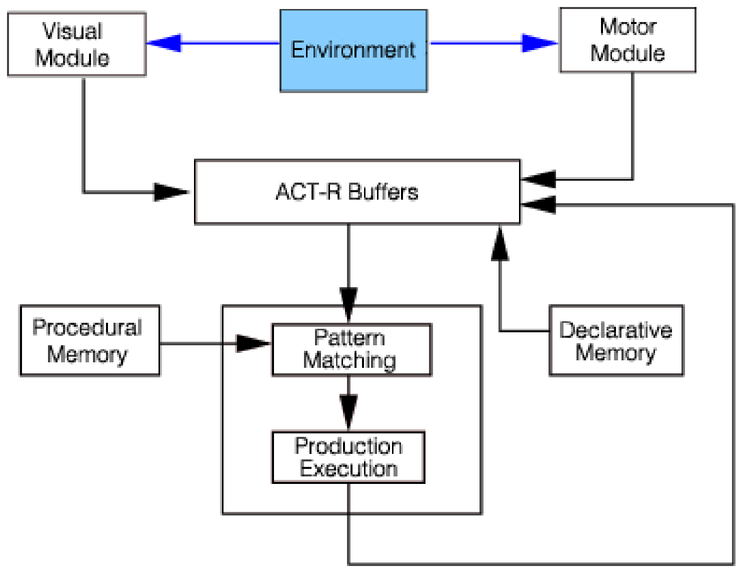

Figure 3: ACT-R module configuration

- Computer-generated forces to provide cognitive agents that inhabit training environments [7].
- Neuropsychology where it has been used to interpret FMRI data [5].

Especially relevant to the subject of this thesis is Boehm-Davis application of ACT-R modeling in trying to improve human performance in an aviation pilot setting; they did this by creating a model based on a cognitive task analysis complemented by empirically collected data from commercial pilots. The end result suggested increased performance of decisions by the pilots [9]. The use of a model based simulation is obvious in safety critical environments such as driving and flying, if a comprehensive model can be built from modeling of relevant cognitive skills involved one could effectively test and evaluate systems quickly and without risk of harm to anyone. The drawback though is similar to that of simulator testing; as it is a simulation, this time of the driver him-/herself, one has to take into account the facts of how it differs from real world use and situations. Through comparison with empirical data this problem can be addressed to some extent. Ritter et al states another advantage of using cognitive models which is that they could easily be developed into agents to help the user with a task, having potential benefits in the driving environment [41].

3.7 Method Theory

3.7.1 Simulator vs. real world testing

In order to evaluate an existing system or a prototype it is important to test it in a driving environment to extract relevant empirical data on for example driving behavior, human-machine interaction, workload or acceptance. This can be done in either a simulator or in a real car in a field study, where both types of tests can be of different degrees of similarity to the real world. When choosing the most appropriate method a choice has to be made between the degree of reality needed and the flexibility and data extraction possibilities, a typical simulator falls in the middle of both these aspects [2].

The advantages of a simulator includes the possibility of controlling most influencing factors on an experiment such as rain and traffic, it also reduces the risk of unforeseen and unwanted traffic situations. Another great advantage is of course that the driver is never in any danger so hazardous situations can be tested [48]. There are however ways to do this in the real world with low danger to the driver, see for example [34] where a real car was driven on the test track but the car the driver would be in danger of hitting was made of light weight plastic foam.

The problem with simulator testing is that the test situation is significantly different than driving a car for real, unless using a simulator with a very high fidelity. In order to be able to draw conclusions from a simulator test and apply those to the real world equivalent, the simulator needs to be validated as well as the situations tested in the simulator. As this is a rather difficult task, caution should be exercised when generalizing the results to real situations.

3.7.2 Measuring workload

NASA-TLX can be used to measure the subjective experience of workload in six domains: mental demand, physical demand, temporal demand, performance, effort and frustration[15]. It has been used in a large number of experiments as a complement to physical measurements or as the sole mean of assessing workload. The advantage of using a self-report measurement lies in the speed and ease of implementation as well as lacking the need of external apparatus, such as heart rate or EEG sensors, having to be attached to the test subject which requires special knowledge in order to be calibrated correctly as well as produces a lot of redundant data.

According to O'Donnel & Eggmeier, problems associated with self-report measures relating to workload include possible confusion of mental and physical load caused by the operator not being able to distinguish between external demand and workload [37]. Other problems shared by all self-report measures are limitations of the inability to introspect as intended, in the workload domain this may cause some operators to rate peak workload while others may rate average workload [47]

3.7.3 Reaction times

Reaction time is important for validation of the simulator as well as the experiment to see if it is, in this respect, possible to generalize the findings to other situations. If the reaction times are found to be very long or very short in comparison to real world driving something in the scenario or the simulator is diverging to much from reality to allow for generalization.

It is of importance to the reaction time if the test is carried out under alerted conditions, where an indictor of some sort is used to alert the driver of the fact that thay are about to encounter a situation where they should break. Unalerted conditions are similar to those presented to a driver in a real driving situation where no warning is given before the situation arise. Longer reaction times are expected under an unalerted condition [42]. Further complicating the research area of reaction times is that different complexities in situations cause different reaction times, the more information that has to be taken into account, the longer the reaction time [22]. The studies adressed here concern reaction time from perception of critical situation to initiated break reaction.

According to Sooebom et al [42], Koppa et al compared reaction times between alerted and unalerted conditions and found reaction times to be an average of 0.60 s under the alerted condition and 0.82 s under unalerted condition [27]. Johansson et Rumar tested reaction time under the unalerted condition and found reaction times to range from 0.66 to 2.00 seconds [23]. These studies were carried out under real driving. Kiefer et al studied brake reaction time in an unalerted setting and found an average reaction time of 1.18 seconds [25]. Lind [31] tested four different types of FCW systems in a similator and found average break reaction times to range from 0.9 seconds to 1.1.

Figure 4: Model of distraction task

4 Method

4.1 Apparatus

A secondary task was used do distract drivers at chosen points during the scenarios. A secondary anticipated effect would also be an increased cognitive load as a result from having to divert attention. In order not to increase this demand to much, which would have an effect to some extent on the results imparted from the driving scenarios, a simple task was used which did not involve demanding analytical cognitive skills.

The secondary task was built in the Flash Air environment from Adobe Inc, the presentation of the stimuli consisted of a 7" display (figure 4) placed in the center stack and connected to the main test computer . At certain points in the scenario the display would give a sound similar to that of a sonar to attract the attention of the test person, this was followed by random numbers between 0 and 100 shown on the display in dark blue on a bright blue background, easily distinguishable, that the driver would have to read and repeat aloud. When not triggered the display was simply black, making color change as well as the sonar sound queues for attention. The trigger for the display was based on the headway distance to a vehicle in front of the driver at a distance of 120 meters.

It was important to place the screen in the center stack for two main reasons; it is part of the driver environment where interruptions in the main task is common when driving without passengers when for example changing radio/cd and other technology often placed in the center stack. It is important to place the visual distraction in such a place that it doesn't give advantages to experienced drivers who have a greater degree of performance in their periferal field of view [14]. Placing the distraction in the center stack low enough that little of the roadway is in the peripheral view ensures this.

The hardware of the warning system used in this study is the FCW implemented in Volvo cars [31] and consists of an array of 12 LEDs paired with a warning sound consisting of three sound pulses. The LED array was placed di-

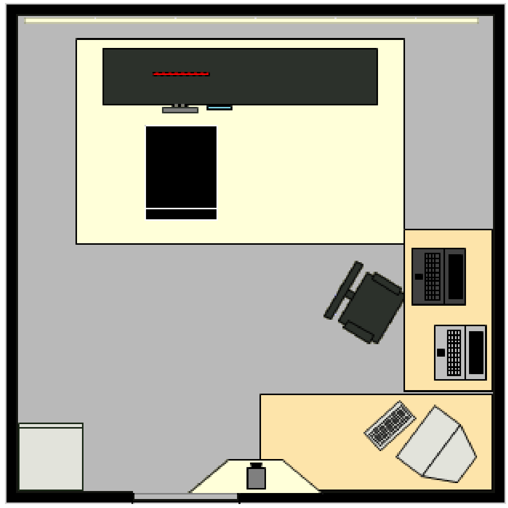

Figure 5: Schematic of simulator room

rectly in the line of sight when looking forward. Unlike Volvo's implementation of the system a reflecting HUD solution was not opted for as the position of the warning system and light conditions made it easily distinguishable to the driver without this in the driving simulator. Introducing a small glass screen would have distorted the projection of the simulated scenario. Warning sound was played back through speakers placed in the ventilation part of the dashboard which is a different sound source than the environmental sounds such as from engine and tires.

The LED array was connected to a BasicX24 microcontroller which in turn communicated with a Flash Air application on a laptop through a serial connection over a serial to USB Bridge. Communication with the Flash Air application was one way, when the threshold value for a warning was reached, the warning would sound and the LED array was triggered in tandem.

The experimental setup (figure 5) consisted of a driving simulator connected to a projector at a resolution of 1280*1024. Connected to the simulator were two laptop computers, one which ran a speedometer from the output of gazelle and the other which ran the secondary task, communication with the simulator and the warning system and also logging of data from the latter. Questionnaires were carried out in Excel before and after testing on the second laptop.

The simulator consisted of two parts, the main STISIM simulator running the simulation and Gazelle which was used to interpret data from STISIM and forward this to the laptops. In effect Gazelle simulated the behavior of events carried out in STISIM.

4.1.1 STISIM and Gazelle

The STISIM Drive simulator, version 2.0, from Systems Technology Inc was used to carry out the tests; it is a simulator fulfilling a requirement of having an easily manipulated scenario building language. This simulator in many different versions is, and has been, used in many research settings at universities around the globe, the version used here is in its most basic form with respect to software capabilities. The drawback to the relative ease of use was that it also imposed some restrictions on what was possible to do with regards to tracking behavior, this was in part solved by the use of the Gazelle simulator, built by Kristoffer Gillenskog and Tobias Åström [20]. Gazelle interprets data from STISIM Drive and passes it on to a desired system by use of TCP/IP protocol. By simulating the behavior of cars in STISIM Drive this enables programming systems such as the warning display system which reacts to events in the scenarios. The system was not able to simulate all desired situations unfortunately which left a constraint on the design of scenarios.

4.2 Independent variables and experimental setup

The experiment was carried out with a between group design. The two groups was presented with scenarios differing in type of situations where the setup in Scenario 1 was prone to giving both false positive alarms and nuisance alarms, while Scenario 2 was prone to nuisance alarms, manifested in the situations where the first group received false positive alarms in addition to the controlled variables of nuisance alarms, alarm 1 and 4. Two of a total of six alarm situations were treated as independent variables, these were situations 3 and 5.

The roadway of both scenarios was the same; it consisted of a 2-file country road with a speed limit of 90 km/h stretching 20 kilometers which gave a completion time of circa 12 minutes depending on speed. During this time the test person would experience 6 alarms from the FCW with a varying degree of correctness which is further explained in the next two sections. At three of these warnings the driver would be distracted by the secondary task, two of which were in an immediate crash situation and one which was not. The intention with the latter type of distraction was to make the scenario less predictable to the driver.

At the critical alarms the warning was triggered at a headway distance of 60 meters which was when the lead vehicle suddenly braked, giving the driver circa 2 seconds reaction time, depending on speed, in which to bring the vehicle to a full stop.

Alarm situation 2 and 6 was designed to be critical situations and reaction times at these were constructed with the intention to be used as the basis for comparison between the two groups. Alarm situation 1 was used to introduce the warning system to the test subject and required no reaction. A further explanation of the different situations used in the scenarios is given below in figures 6 to 9.

4.2.1 Explanation of scenario situations

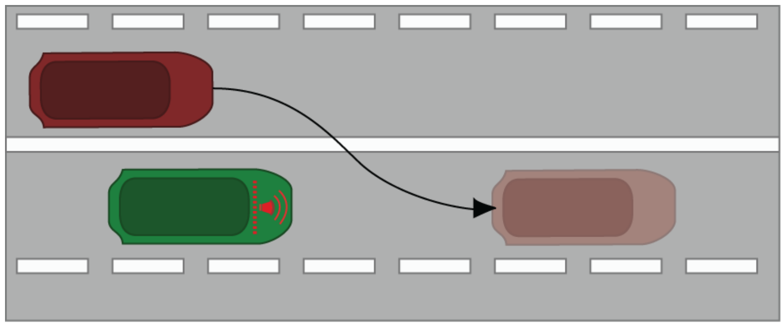

Figure 6: Situation type A –Nuisance warning. Car overtaking illegally
A car approached the driver rapidly from behind and made an illegal overtaking across the solid middle stripe. The FCW was triggered when the overtaking car pulled back into the lane close to the driver. This situation did not warrant any action of avoidance on behalf of the driver.

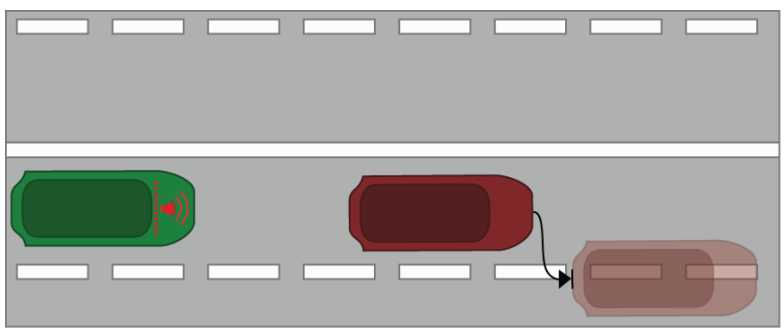

Figure 7: Situation type B – Nuisance warning. Lead vehicle suddenly turns and stops by the side
At the same distance as in the critical situation, 60 meters, the lead vehicle abruptly pulled over to the side of the road. Leading up to this situation the lead vehicle kept a speed of -5km/h relative to the speed of the driver's vehicle in order to ensure reaching the threshold distance of 60 meters causing the sudden breaking. At 120 meters the secondary task was triggered. The situation did not require any reaction in order to avoid a collision with the lead vehicle but does not have a large safety margin.

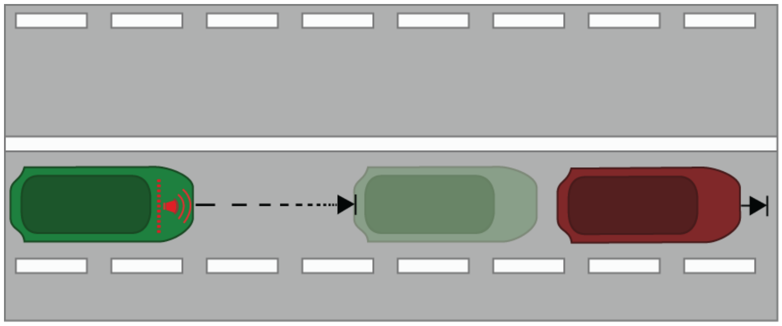

Figure 8: Situation Type C – Critical warning. Lead vehicle suddenly breaks to a halt
This situation occurred when the lead vehicle suddenly braked for no apparent reason at a distance of 60 meters from the driver. This gave the driver a reaction time of circa 2 seconds depending on speed with the only legal reaction being to brake. Leading up to this situation the lead vehicle kept a speed of -5km/h relative to the speed of the driver's vehicle in order to ensure reaching the threshold distance of 60 meters causing the sudden breaking. The secondary task was triggered at a distance of 120 meters.

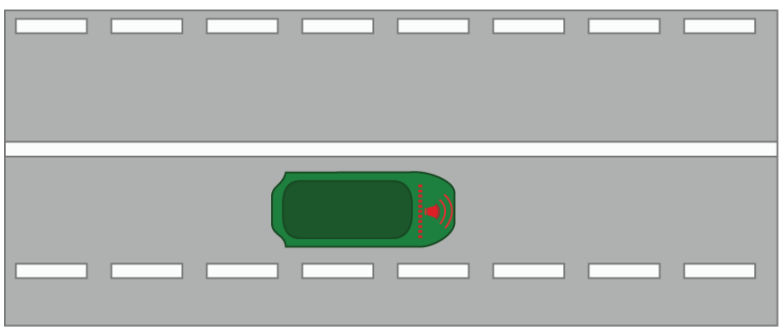

Figure 9: Situation type D – False Positive warning. No visible reason
In this situation the FCW was triggered by no apparent reason and thus did not warrant a reaction from the driver.

4.2.2 Scenario structure

The Nuisance Prone scenario and the False Positive Prone scenario as is being shown in figures 10 and 11 differs in situations 3 and 5. The small diversity of situations presented to the driver was limited by the capacity of the simulator; this explains why situations repeat, for example situation type A was used for introduction of the FCW in the first situation as well as in the capacity of a nuisance alarm in situation 5 of the nuisance prone scenario. The situations were spaced 2 minutes apart, about +- 20 seconds. Defining the warnings as

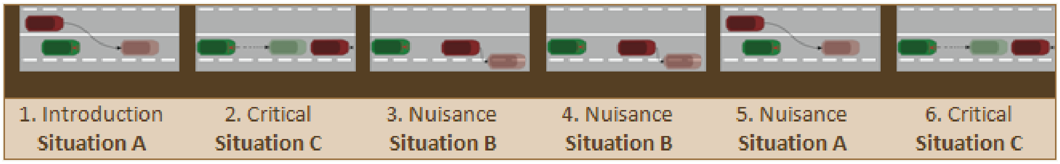

Figure 10: Scenario 1, Nuisance Prone (N)

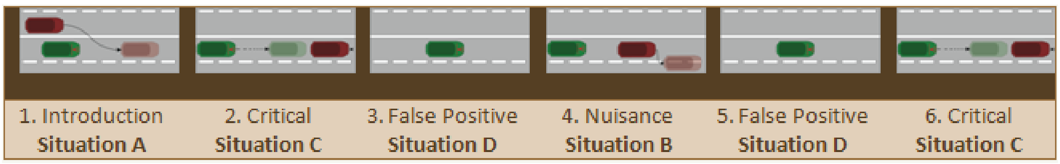

Figure 11: Scenario 2, False Positive Prone (FP)

unalerted or alerted poses a problem as in one sense they were alerted as the participants were aware that the focus of the study was warning systems in cars, at the same time the driver was not alerted as such before the situations and thus is to be seen as unalerted warnings.

4.3 Dependent variables

To elicit the difference of effect on the test subjects, the reaction time was logged with regard to accelerator release and brake initiation as an objective measurement. Reaction time was measured from initiation of the frontal collision warning to a breaking reaction, the FCW was triggered at the same time as the vehicle in the headway of the driver changed behavior. Subjective experience of workload was elicited through a NASA-TLX questionnaire (appendix D) followed by a questionnaire focused on trust to the system as well as number of estimated number of warnings experienced (appendix D), the questionnaire was randomized in four variations in an effort to cancel out the order of questions affecting the results. During the simulation run, a DV camera was used to document where the test person was looking in order to discern whether he or she was looking at the distraction task at the time of warning. Initially, an eye-tracking system was tried to automate this but due to time consumption when calibrating it for each participant as well as problems with synchronizing the log from the eye-tracker with the other logs, it was decided to go with the more time efficient, during tests, DV camera solution.

4.4 Test subjects

29 test subjects participated in the study; 8 females and 21 males ranging in age between 21 and 50. All participants were licensed drivers since at least 2 years and had normal or corrected vision. Five participants had previous

Total averages *(n=21)*	Age	Driving Regularity	Time Licensed	Km/month
Mean	30.7	2.1	3.3	83.1
Median	27	3	2	60
Std dev	11.9	1.1	1	72.6

Table 2: User statistics

N Prone *(n=9)*	Age	Driving Regularity	Time Licensed	Km/Month
Mean	33.1	1.9	3.1	85.6
Median	26	2	3	80
Std dev	10.8	1.1	1.1	65.8

FP Prone *(n=12)*	Age	Driving Regularity	Time Licensed	Km/Month
Mean	28.9	2.3	3.5	81.3
Median	28	2.5	3.5	50
Std dev	12.8	1.2	1	80.2

Table 3: Comparison of test groups

experience of warning systems, specifically Frontal Collision Warning and Blind Spot Warning.

Invitations was posted around the university campus and spread through emails to Volvo and SAFER with an address where participants could respond. Participants were informed of the domain of the study, i.e. warning systems in cars in the invitation; student participants received a cinema ticket upon completion of the test. 10 participants were employed by Volvo and care was taken so that they would not have been part of the development of the system. Student and Volvo employees were divided equally between the two groups. Sex and age was not taken into consideration in the choice of participants as the amount of respondents didn't allow for this, participants were assigned scenarios randomly. Driving regularity was divided into 5 categories (1: Daily, 2: Each week, 3: Each Month, 4: Each Year, 5: Never), if someone answered "never" they would be excluded. Time licensed was also divided into 5 categories (1: 0-2 years, 2: 2-5 years, 3: 5-10 years, 4: 10-20 years,5: >29 years) as in the previous question if the answer was "0-2 years" they would be excluded. No participant fell into either of these categories. Detailed statistics on the test subjects that correctly completed the test can be found in table 2. The difference between the two groups can be found in table 3.

5 Results

The results in this section are based on the 21 test subjects shown in the previous chapter. 29 participants carried out the tests but due to one case of nausea, three cases of system malfunctioning and two participants crashing without initiating a breaking maneuver which had the result of reaction data being recorded. Two participants were removed as their reaction times where considered outliers on the basis of having negative reaction time, i.e. initiated breaking maneuver before any situation had arisen, the data of these 8 test subjects were thusly discarded

Detailed tables of data showing Average, Median, Standard Deviation and standard error for the graphs presented in this chapter can be found in Appendix A.

5.1 Reaction times

Analysis of a correlation matrix between the sets of objective data was carried out to find diverging correlations warranting further analysis which is presented here.

Figure 12 shows the mean results of reaction time measured from alarm onset to initiating breaking maneuver in the two scenarios. A paired t-test analysis of the reaction time regarding the first critical warning (C1) between the two groups showed no statistical significance as expected ($p > 0.05$). The same analysis carried out on the second critical warning (C2) between the scenarios also fail to show any statistical significance ($p > 0.05$). Analysis of the standard error of mean supports this analysis. No significant difference was found when comparing between age groups (< 30 vs > 30) or between males and females.

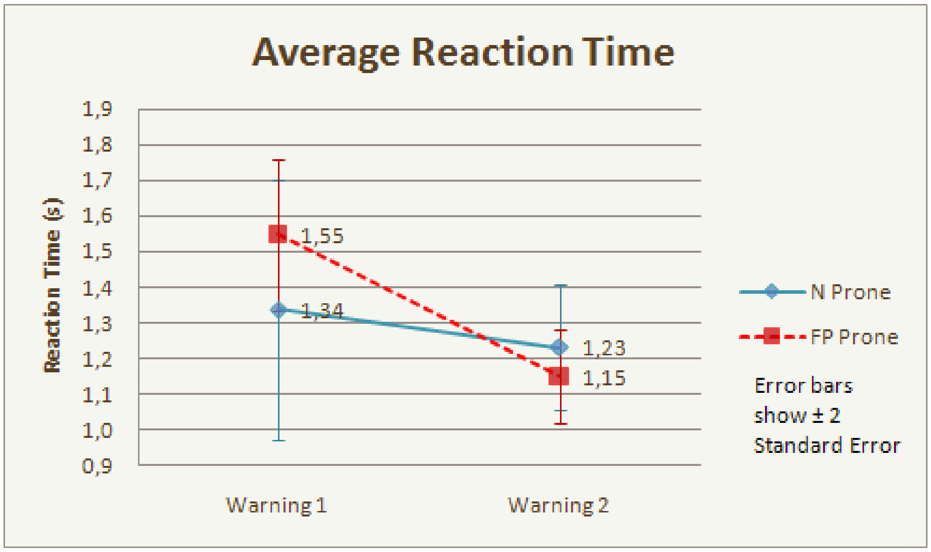

Figure 12: Average reaction time

In order to see if the similarity of the two groups is caused by the slightly skewed factor of driving experience between the groups being an influencing factor, a comparison was made between frequent drivers (Driving Regularity category 1 and 2) and infrequent drivers (category 3 and 4). Time licensed was also interesting due to the same reasons and was divided into moderate time licensed (Time Licensed category 2 and 3) and long time licensed (category 4 and 5). Since there is a possibility of having had a drivers' license for a long time without driving regularly a coefficient of experience was calculated by dividing the mean "category value" of time licensed with the same of driving regularity. Time licensed was reduced by 1 to create scales with equal distributions of intervals. Drivers with a coefficient > 1.0 was considered experienced drivers which implies for example that a participant who has had a license for 2-5 years and drive daily belong to the less experienced group with a coefficient of 1.0 while someone who has been licensed for 10-20 years and drive weekly belong to the experienced group with a coefficient of 1.5. The two groups were then compared with each other with regard to reaction time, the results can be found in figures 13 and 14.

Paired t-test comparison between the two groups reveals no significant difference between the groups in either scenario ($p>0.05$), this is confirmed by error bar analysis. The origin of the difference between the two scenario groups is revealed though when looking at the influence of experience on reaction time, the low degrees of freedom in each however restricts the statistical power of the test.

The lower mean values of reaction time at C2 compared to C1 imply that some sort of learning behavior is possible to have taken place.

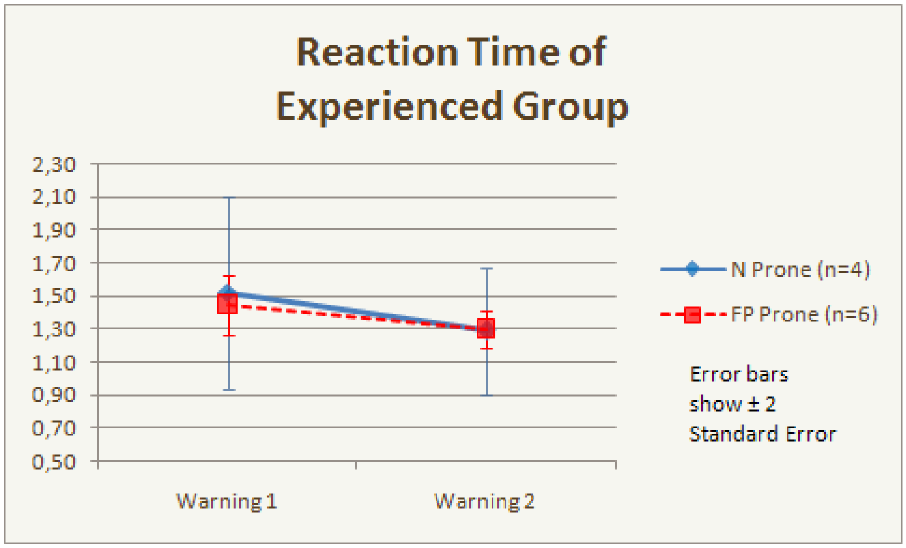

Figure 13: Average reaction times, experienced drivers

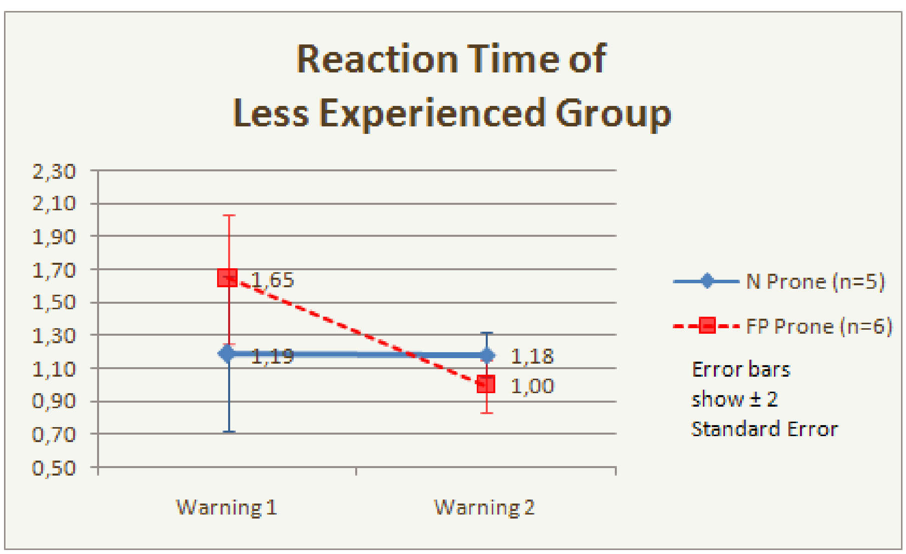

Figure 14: Average reaction times, less experienced drivers

5.2 NASA-TLX

Figure 15 show the mean results of the NASA-TLX questionnaire. There does not seem to be a large difference between the groups when looking at the mean values, the error bars reveal that the spread of answers is rather high, making it difficult to draw any conclusions.. These findings are the same both for the total mean of the parameters as well as for the individual parameters.

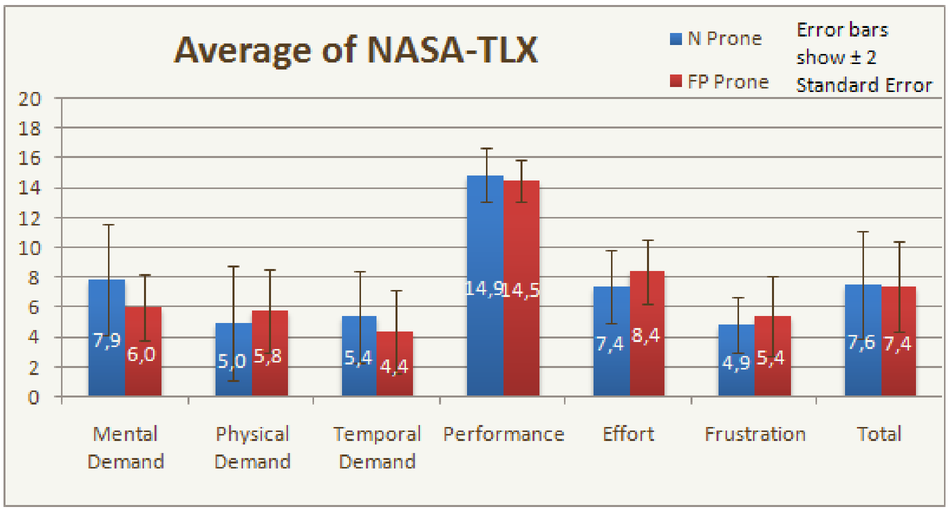

Figure 15: NASA-TLX results

5.3 Trust

In order to assess a level of overall trust in the system for comparison between the two scenarios, negative statements were reversed. A structured version of the original questionnaire can be found in appendix D.

Three answers were left blank by respondents, probably by mistake. These were for the N Prone scenario question 14, for the FP Prone scenario they were question 21 and 23, one respondent in each case, these questions were omitted for this analysis resulting in a lower number of respondents for those questions but the rest of the answers were factored into the analysis. There was no reason to believe that they were left out of privacy issues or other reasons than missing them in the answering form as their nature where similar to other questions that where answered.

To assess a mean overall rating of how beneficial the system was experienced to be by the test subjects, all answers of the reversed questionnaire for each scenario was summed up and compared. Two questions were left out of this analysis as they did not pertain to a negative/positive nature (questions 18 and 19).

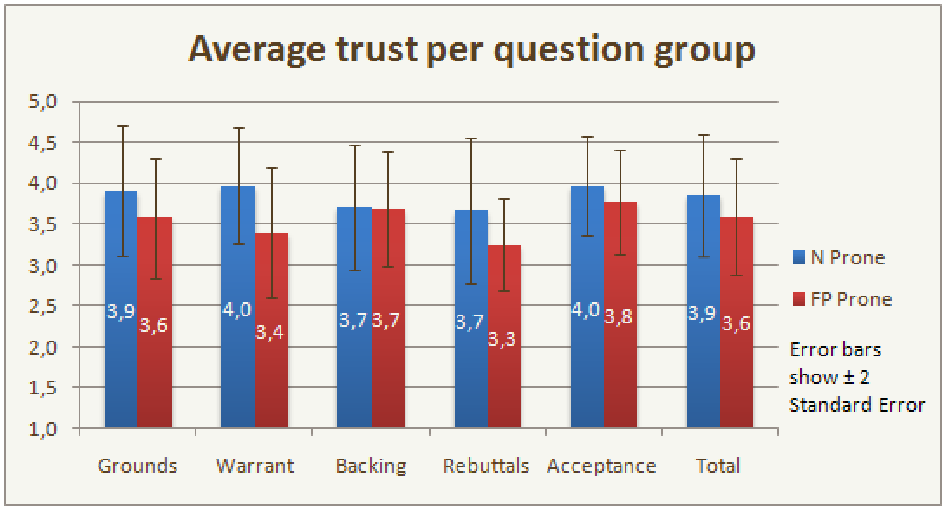

Figure 16: Questionnaire results

A comparison between questions regarding, Grounds, Warrant, Backing and Rebuttals was carried out between the two groups. This was used to attempt an extrapolation of the argument part of the APT framework as well as a subjective measurement of overall trust. The detailed results can be found in figure 16. The results were compared by standard error of mean but no definitive difference was found, error bars show a wide spread.

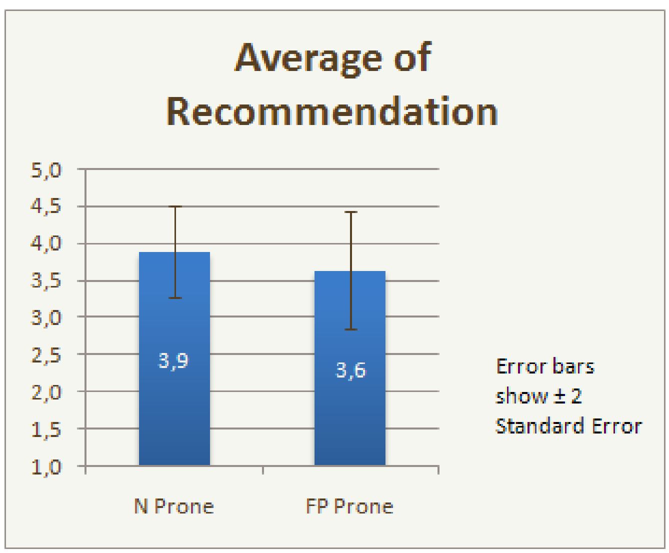

Figure 17: Recommendation

One question (23) regarding whether the test subjects would recommend the system to others was used to assess overall experience with the system (figure 17). No significant difference was found between the two groups when comparing standard error of means.

5.4 Subjective estimate of number of warnings

A subjective estimate of number of warnings experienced by the test subjects, as well as the estimated number of unnecessary warnings, hold interest to the parameters part of the APT framework, showing the difference in calibration of the participants' trust models. It is also important in order to see whether the nuisance warnings were understood as closer correct warnings or false, if all nuisance warnings would be considered as false positive warnings the estimated number of incomprehensible warnings over both scenarios would be the same. Comparison between the two scenarios by means of standard error indicates a significant difference between the number of estimated incomprehensible warnings.

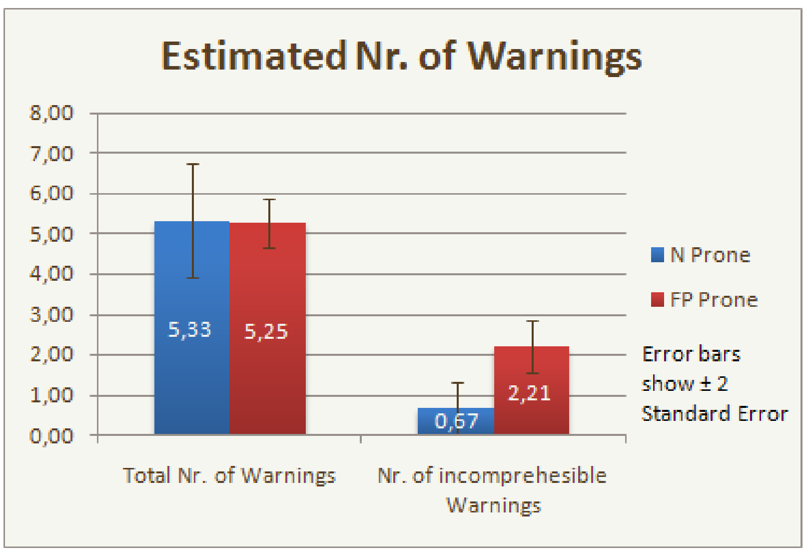

Figure 18: Estimated number of warnings

5.5 Observed behavior

The distraction success rate of the secondary task can be found in the table 4. Distraction success is here defined as the driver having focus on the distraction task at the moment of warning; all participants successfully complied with the distraction task indicating that they allocated cognitive resources towards it. The much higher rate of distraction at the final warning in the FP prone scenario, when compared to the other situations, is interesting and has to be explained in conjunction with other findings.

N Prone	C1	C2
Percentage	54.5%	45.4%
FP Prone	C1	C2
Percentage	58.3%	75.0%

Table 4: Distraction success rate

The use of a DV-camera to monitor distraction also enabled other observations; most prominently it was observed that in the FP prone scenario, half of the participants continued to count the numbers in the distraction task after the final critical warning while none of the participants of the N Prone scenario did this.

6 Discussion

6.1 Performance

When looking at average reaction times, no significant difference could be found between the two groups which imply that the effect of nuisance warnings compared to that of false positive warnings is small. The origin of this result needs to be analyzed with some support; it may be that all participants regarded the independent variables as the same type of warning, if so most probably as false positive warnings according to the definitions presented in this thesis, or it may simply be that there arises no performance difference between the two types of warnings. If we look at the results from the estimated warnings (figure 18) we find that the two groups estimated different amounts of incomprehensible warnings in the intended fashion, it therefore seems likely to conclude that the lack of performance difference indicates that there was no actual performance difference between the two types of warnings.

There seems to be a tendency, however not statistically significant, of learning behavior which was found to originate from inexperienced drivers. This should be interpreted with caution as the statistical power is very low and may be caused by randomness in the data but may be worth noting for future studies. A possible explanation could be that more experienced drivers adapt to a new driving situation faster and therefore are less influenced by the rather artificial setting of the simulator.

When looking at the rate of successful distraction with respect to the secondary task it is worth noting that the FP Prone scenario group had a much higher success rate. A possible explanation is that the drivers increased their vigilance of events around them due to being presented with a more unstable system and thus was able to better carry out the secondary task; this should increase the workload as the warning then present as a distraction but no supportive evidence of this was found. Also affecting the secondary task could be that the FP warnings reduce compliance to the system and therefore may free up cognitive resources to focus on other tasks while the drivers in the nuisance prone scenario had more focus on the warning system. It is interesting to look at this in conjunction with the observation that the FP Prone scenario group continued to carry out the secondary task after the last situation was dealt with which could be a signifier of increased focus on the execution of the scenario tasks.

6.2 Trust

No significant differences were found in any domain of the questionnaire though some tendencies can be seen of lower scores in all areas except Backing by the FP Prone group. These results are in line with what was expected from the theories behind the questionnaire however as the statistical support is very weak no conclusions can be drawn.

There was no significant difference between the two scenarios regarding recommendation of the system to others. Considering that the test persons estimated the number of incomprehensible warnings rather close to their actual numbers, this seems to have little bearing on their appreciation of the system. Previous studies addressed in this thesis show that operators will accept and follow a faulty system and these findings seem to be in line with this research, at least in the acceptance part. As reasoned previously, drivers with the FP Prone FCW may have disregarded its warnings, but it is then surprising that they would recommend it as much as those using the N Prone FCW. Looking at the score given to the question we see that there does not seem to have been a strong preference for the system in either group though both were on the positive side, it is possible that the answer given to this question is slightly inflated out of courtesy and the cause for the similar results lie in the fact that both groups had a cautionary approach to relying on and complying with the system.

6.3 NASA-TLX

The results of the NASA-TLX measurements does not warrant any conclusions to be drawn, the differences of results between the two scenarios are quite small and thus seem to indicate that workload is not influenced by the type of warnings experienced. One reason for this may be that when facing FP warnings the driver disregards the system and thus frees up workload as previously discussed. It should be noted that the power of the NASA-TLX measurement is lowered by the lack of any physical measurement.

6.4 Method discussion

Here it will be discussed how the methods may have influenced the results, whether they were found to be useful and whether they allow for conclusions to be drawn.

6.4.1 Trust questionnaire

The questionnaires used in this study was created from scratch with influence from studies of trust and heavily based upon the APT framework [12]. There is always a problem with creating a new questionnaire and subsequently drawing conclusions from it as it contains no element of validation; an established questionnaire that has been used widely has a higher likelihood of being validated. This may lead to conclusions becoming over generalized as well as differences found between groups become hidden by the structure of the questionnaire, restricting its power as a research tool. Efforts of finding a validated general trust questionnaire were made but did not result in any useful findings. In order to discern how trust works, in this case according to the APT framework, a separate questionnaire would then have to be constructed to be used as a compliment.

It follows that the results of the trust questionnaire used should be seen as a prospective way of discerning levels and workings of trust rather than taken at face value. Relating the results to theory and correlating it with results from other data from the study was an instrumental part of the analysis of the trust questionnaire.

6.4.2 NASA-TLX

NASA-TLX was used as the sole measurement of workload in this study, as workload is a complex thing it is preferable to use this method in conjunction with a physical measurement such as EEG or measurement of glance time. These things were not possible unfortunately as the time and knowledge required was not at hand. Efforts of incorporating eye tracking apparatus was tried but resulted in more problems than advantages due to technical problems, this could however have been used for measurement of glance time as a signifier for level of workload.

6.4.3 Simulator validation

The reaction times recorded from this study seem to lie within reaction times from the naturalistic studies that has been adressed in this thesis; this constitutes a validation of the simulator setup and scenario design used here as a research tool which allows for conclusions to be useful in a larger perspective.

7 Conclusions

The findings of this thesis provide an insight into the effect of nuisance warnings and how they relate to false positive warnings. The results indicate that the nuisance alarms had the same effect, or lack thereof as false positive warnings even though they were, in contrast to false positive warnings, identified as having a reason and thus were understood. Lack in difference of mental workload as well as trust in the system support these conclusions. The hypothesizes and research questions are answered below.

Null hypothesis: There will be no significant difference in reaction times between the two test groups as well as no difference in workload or trust in the system

The findings in this thesis failed to reject the null hypotheses in any regard thus concluding that, under the tested circumstances, no difference could be shown between false and nuisance warnings as they are defined here. The alternate hypothesis may therefore be discarded.

What are the effects of false alarms versus nuisance alarms in a driving environment?

As mentioned no difference in the objective measurements was found, there was however a difference in the behaviour of the test subjects between the two settings. In the false alarm setting there was a significantly higher degree of compliance with the distraction task at the final critical warning and they also continued to comply with it after they had stopped the vehicle. In effect the nuiscance group considered stopping the vehicle at the final warning as the end of the scenario (they were aware of the estimated time the test would take) while the false positive group considered the end of the secondary task as the end of the scenario.

How will the driver's subjective estimate of trust differ between the different scenarios?

No significant difference could be found between the scenarios in any area of the drivers conceptual interpretation of the system when analysing the novel trust questionnaire constructed for this thesis.

8 Future Work

The research in this thesis poses a starting point for looking at aspects of nuisance alarms that is in need of extrapolation. Next proposed step in this line of research would be to compare nuisance warnings with correct warnings in an attempt to validate the findings in this thesis and also to complete the full picture of implications of nuisance warnings. Also a longitudinal study showing the long term effects of nuisance alarms is proposed as this may differ from the type of study carried out here.

The data found here could be part of building a model in the ACT-R framework, specifically modules governing decision processes and stimuli reaction. The development of an accurate model capable of interaction could enable great a leap forward in the research on vehicle safety.

References

[1] Injuries in the european union: Statistics summary 2003-2005 vienna, october 2007 isbn: 978-3-7070-0081-8.

[2] Alferdinck, J. W. A. M., Hogema, J. H., Theeuwes, J., and Van der Horst, A. R. A. Methods for measuring perception and driving behaviour in darkness and fog. *Proceedings of the First international congress "Vehicle & Infrastructure Safety Improvement in Adverse Conditions and Night Driving" VISION* (24 and 25 September 2002).

[3] Anderson, J., Bothell, D., and Byrne, M. An intergrated theory of mind. *Psychological review 111(4)* (2004), 1036–1060.

[4] Anderson, J., and Lebiere, C. *The Atomic Components of Thought*. Lawrence Erlbaum, 1998.

[5] Anderson, J. R., Albert, M., and Fincham, J. Tracing problem solving in real time: fmri analysis of the subject-paced tower of hanoi. *Journal of Cognitive Neuroscience 17(8)* (2005), 1261–2174.

[6] Besnard, D., Greathead, D., and Baxter, G. When mental models go wrong: Co-occurrences in dynamic, critical systems. *International Journal Of Human Computer studies 60* (2004), 117–128.

[7] Best, B., and Lebiere, C. Cognitive agents interacting in real and virtual worlds. In *Cognition and Multi-Agent Interaction: From Cognitive Modeling to Social Simulation*. Cambridge University Press, 2006.

[8] Bliss, J. P., and Acton, S. A. Alarm mistrust in automobiles: How collision alarm reliability affects driving. *Applied Ergonomics 34(6)* (2003), 499–509.

[9] Boehm-Davis, D., Holt, R., Diez, M., and Hansberger, J. Developing and validating cockpit interventions based on cognitive modeling. *Proceedings og the Twenty-Fourth Annual Conference of the Cognitive Science Society* (2002), 27.

[10] Brookhuis, K., De Waard, D., and Janssen, W. Behavioural impacts of advanced driver assistance systems: An overview. *EJTIR, 1 3* (2001), 245–253.

[11] Budiu, R. About act-r. Retrieved 18/1 2009 from http://act-r.psy.cmu.edu/about/.

[12] Cohen, M. S., Parasuraman, R., and Freeman, J. T. Trust in decision aids: A model and its training implications. *Command and Control Research and Technology Symp. Monterey CA* (1998), 1–37.

[13] Congress, N. The automated highway system: An idea whose time has come. *Public Roads 58(1)* (1994).

[14] Crundall, D., Underwood, G., and Chapman, P. Driving experience and the functional field of view. *Perception 28* (1999), 1075–1087.

[15] De Waard, D. *The Measurement of Drivers' Mental Workload*. PhD thesis, University of Groningen, Haren, The Netherlands, University of Groningen Traffic Research Centre VSC,, 1996.

[16] Dixon, S. R., Wickens, C. D., and McCarley, J. How do automation false alarms and misses affect operator compliance and reliance? *Proceedings Of The Human Factors And Ergonomics Society, 50th Annual Meeting* (2006), 25–29.

[17] Dragutinovic, N., Brookhuis, K., Hagenzeiker, M., and Marchau, V. Behavioural effects of advanced cruise control use: A meta-analytic approach. *EJITR, 5 4* (2005), 267–280.

[18] Enriquez, M., and MacLean, K. Impact of haptic warning signal reliability in a time-and-safety-critical task, haptic interfaces for virtual environment and teleoperator systems. *HAPTICS '04. Proceedings. 12th International Symposium* (2004), 407– 414.

[19] Gambetta, D. *Trust: Making and Breaking Cooperative Relations*. Department of Sociology, University of Oxford, 2000, ch. 13. Can We Trust Trust?, pp. 213–237.

[20] Gillenskog, K., and Aströom, T. Gazelle - kommuniaktion med bilsimulator. Master's thesis, Department of Computer Science, Chalmers University of Technology, 2008.

[21] Headly, P. Esc as a baseline for active safety. *Proceedings of 19th ESV, NHTSA Paper No. 05-0332* (2005).

[22] Hutchinson, T. Tailgating. Tech. rep., Centre for Automotive Safety Research, The University of Adelaide, AUSTRALIA, 2008.

[23] Johansson, G., and Rumar, K. Drivers brake reaction times. *Human Factors 13* (1971), 23.27.

[24] Kazi, T., Stanton, N., Walker, H., and Young, M. Designer driving: Drivers conceptual models and level of trust in adaptive cruise control. *International Journal of Vehicle Design 45(3)* (2007), 339–360.

[25] Kiefer, R. J., LeBlanc, D., Palmer, M., Salinger, J., Deering, R., and Shulman, M. Camp report: Development and validation of functional definitions and evaluation procedures for collision warning/avoidance systems. Tech. rep., U.S. Department of Transportation, NHTSA, 1999.

[26] Kircher, K. Driver distraction: a review of the literature. Tech. rep., VTI (Swedish National Road and Traffic Research Institute), 2007.

[27] KOPPA, R., FAMBRO, D., AND ZIMMER, R. Measuring driver performance in braking maneuvers. *Transportation Research Record: Journal of the Transportation Research Board 1550* (1996), 8–15.

[28] LEE, J. D., HOFFMAN, J., AND HAYES, E. Collision warning design to mitigate driver distraction. *Letters of the CHI 6(1)* (2004), 65–72.

[29] LEE, J. D., AND SEE, K. A. Trust in automation: Designing for appropriate reliance. *Human Factors 46* (2004), 50–80.

[30] LEWIS, M. W., MILSON, R., AND ANDERSON, J. R. The teacher's apprentice: Designing an intelligent authoring system fro high school mathematics. In *Artificial Intelligence and Instruction*. Addison-Wesley, 1987.

[31] LIND, H. An efficient visual forward collision warning display for vehicles. *SAE World Congress & Exhibition* (2007).

[32] LINDGREN, A., AND CHEN, F. State of the art analysis: An overview of advanced driver assistance systems (adas) and possible human factors issues. In *Human Factors and Economic Aspects on Safety. Swedish Network for Human Factors Conference* (1999).

[33] LIU, Y.-C. Comparative study of the effects of auditory, visual and multimodality displays on drivers performance in advanced traveller information systems. *Ergonomics 44(4)* (2002), 425–442.

[34] MCGEHEE, D., MAZZAE, E., AND BALDWIN, S. Driver reaction time in crash avoidance research: Validation of a driving simulator study on a test track. *Human Factors and Ergonomics Society Annual Meeting Proceedings 3(4)* (2000), 320–323.

[35] MEYER, J. Effects of warning validity and proximity on responses to warnings. *Human Factors 43(4)* (2001), 563–572.

[36] MUIR, B. Trust in automation : 1. theoretical issues in the study of trust and human intervention in automated systems. *Ergonomics 37* (1994), 1905–1922.

[37] OŽDONNELL, R., AND EGGEMEIER, F. Workload assessment methodology. *Handbook of Perception and Human Performance 2, Cognitive Processes and Performance* (1986), 42/1–42/49.

[38] PARASURAMAN, R., AND RILEY, V. Humans and automation: Use, misuse, disuse, abuse. *Human Factors 39* (1997), 230–253.

[39] PASHLER, H. Dual-task interference in simple tasks: Data and theory. *Psychological Bulletin 116(2)* (1994), 220–244.

[40] REEVES, B., AND NASS, C. *The Media Equation: How People Treat Computers, Television, and New Media Like Real People and Places*. Cambridge University Press, 1996.

[41] Ritter, F., Baxter, G., Jones, G., and Young, R. User interface evaluation: How cognitive models can help. *Human-computer interaction in the new millenium* (2001), 125–147.

[42] Sooebeom, L., Myungsoon, C., Youngiyun, L., and Wonchul, K. A study of driver's perception response time on the three dimensional alignment with virtual real method. *Journal of the Eastern Asia Society for Transportation Studies 5* (2003).

[43] Sorkin, R., Kantowitz, B., and Kantowitz, S. Likelihood alarm displays. *Human Factors 30(4)* (1988), 445–459.

[44] St. Amant, R., Horton, T., and Ritter, F. Model-based evaluation of cell phone menu interaction. *Proceedings of the CHI'04 Conference on Human Factors in Computer Systems* (2004), 343–350.

[45] Tijerina, L., Johnston, S., Parmer, E., Pham, H. A., Winterbottom, M., and Barickman, F. Preliminary studies in haptic displays for rear end collision avoidance system and adaptive cruise control system applications. Tech. rep., U.S. Department of Transportation, 2000.

[46] Toulmin, S. *The uses of argument*. Cambridge, UK: Cambridge University Press, 1958.

[47] Vidulich, M., and Tsang, P. Techniques of subjective workload assessment: a comparison of swat and the nasa-bipolar methods. *Ergonomics 29* (1986), 1385–1398.

[48] Weiler, J., Bloomfeld, J., Woodworth, G., Grant, A., Layton, T., Brown, T., McKenzie, D., Baker, T., and Watson, G. Effects of fexofenadine, diphenhydramine, and alcohol on driving performance a randomized, placebo-controlled trial in the iowa driving simulator. Tech. rep., American College of Physicians, American Society of Internal Medicine, 2000.

[49] Wickens, C., and Dixon, S. Is there a magic number 7 (to the minus 1)? the benefits of imperfect diagnostic automation: A synthesis of the literature. Tech. rep., University of Illinois, Aviation Human Factors Division., 2005.

[50] Wickens, C., Dixon, S., Goh, J., and Hammer, B. Pilot dependence on imperfect diagnostic automation in simulated uav flights: an attentional visual scanning analysis. Tech. rep., University of Illinois, Aviation Human Factors Division, 2005.

[51] Wiese, E., and Lee, J. Auditory alerts for in-vehicle information systems: The effects of temporal confliict and sound parameters on driver attitudes and performance. *Ergonomics 47(9)* (2004), 965–986.

[52] Xiao, Y., and Seagull, F. An analysis of problems with auditory alarms: Defining the roles of alarms in process monitoring tasks. *Proceedings Of The Human Factors And Ergonomics Society, 43rd Annual Meeting* (1999), 256–260.

[53] Young, M., and Stanton, N. Malleable attentional resources theory: A new explanation for the effects of mental underload on performance. *Human Factors 44(3)* (2002), 365–375.

[54] Young, M., Stanton, N., and Harris, D. Driving automation: Learning from aviation about design philosophies. *International Journal of Vehicle Design 45(3)* (2007), 323–338.

A Reaction time and questionnaire data

	Reaction Time: C1	Reaction Time: C2	Reaction Decrease	Sex	Age	License category	Regularity category	km/month
1	1,63	1,29	0,34	F	23	2	2	100
2	1,44	0,98	0,46	M	47	5	1	167
3	2,37	1,85	0,52	F	26	3	1	80
4	0,72	1,10	-0,38	M	26	3	4	5
5	0,52	1,29	-0,77	F	50	2	2	60
6	1,11	1,21	-0,10	F	37	4	1	50
7	1,58	0,95	0,63	M	23	2	3	8
8	1,49	1,29	0,20	M	25	3	2	100
9	1,17	1,12	0,05	M	41	4	1	200

Table 5: Reaction Times and Test subject data, Nuisance Prone

	Reaction Time: C1	Reaction Time: C2	Reaction Decrease	Sex	Age	License category	Regularity category	km/month
1	1,45	1,12	0,33	M	29	3	3	3
2	1,78	1,38	0,40	M	44	5	1	250
3	1,36	1,22	0,14	F	22	4	2	5
4	1,63	1,39	0,24	M	24	3	1	150
5	1,27	0,93	0,34	F	31	4	3	50
6	1,44	1,42	0,02	F	32	4	2	40
7	1,26	1,05	0,21	M	55	5	1	200
8	2,15	1,19	0,96	M	27	3	4	5
9	1,22	1,35	-0,13	M	33	4	1	83
10	1,12	0,96	0,16	M	23	2	3	100
11	2,30	0,66	1,64	M	21	2	3	50
12	1,60	1,13	0,47	M	26	3	4	40

Table 6: Reaction Times and Test subject data, False Positive Prone

N Prone	C1	C2	Decrease
Average	1.52	1.29	0.23
Median	1.31	1.16	0.25
Standard deviation	0.58	0.39	0.30
Standard error of means	0,15	0,10	0,08
FP Prone	C1	C2	Decrease
Average	1.45	1.30	0.15
Median	1.40	1.37	0.18
Standard deviation	0.22	0.14	0.18
Standard error of means	0,04	0,03	0,04

Table 7: Reaction Times, experienced drivers

N Prone	C1	C2	Decrease
Average	1.19	1.18	0.00
Median	1.49	1.29	0.20
Standard deviation	0.53	0.15	0.57
Standard error of mean	0,12	0,03	0,13
FP Prone	C1	C2	Decrease
Average	1.65	1.00	0.65
Median	1.52	1.04	0.40
Standard deviation	0.48	0.19	0.56
Standard error of mean	0,10	0,04	0,11

Table 8: Reaction Times, less experienced drivers

N Prone	Mental	Physical	Time	Performance	Effort	Frustration	**Total**
Average	7,88	5,00	5,38	14,88	7,38	4,88	7,56
Median	7	2	4.5	16.5	6	4	6
Standard deviation	5,57	5,76	4,69	2,70	3,62	2,75	5,41
Standard error of mean	1,86	1,92	1,56	0,90	1,21	0,92	1,80
FP Prone	Mental	Physical	Time	Performance	Effort	Frustration	**Total**
Average	6,00	5,75	4,42	14,50	8,42	5,42	7,42
Median	5,00	3,50	2,50	15,00	7,50	4,00	6,50
Standard deviation	3,81	4,77	4,85	2,43	3,75	4,62	5,24
Standard error of mean	1,10	1,38	1,40	0,70	1,08	1,33	1,51

Table 9: NASA-TLX results

N Prone	Grounds	Warrant	Backing	Rebuttals	Acceptance	**Total**
Average	3,90	3,97	3,71	3,67	3,97	3,86
Median	4	4	4	3	4	4
Standard deviation	1,20	1,06	1,15	1,32	0,91	1,11
Standard error of mean	0,40	0,35	0,38	0,44	0,30	0,37
FP Prone	Grounds	Warrant	Backing	Rebuttals	Acceptance	**Total**
Average	3,57	3,40	3,68	3,25	3,77	3,59
Median	4	4	4	3	4	4
Standard deviation	1,25	1,38	1,21	0,97	1,11	1,23
Standard error of mean	0,36	0,40	0,35	0,28	0,32	0,36

Table 10: Trust questionnaire results

Scenario	**N Prone**	**FP Prone**
Average	3,89	3,64
Median	4,00	4,00
Standard deviation	0,93	1,36
Standard error of mean	0,31	0.39

Table 11: Recommendation of the system

N Prone	Total warnings	Incomprehensible
Average	5,33	0,67
Median	5,00	0,00
Standard deviation	2,12	1,00
Standard error of mean	0.71	0.33
FP Prone	Total warnings	Incomprehensible
Average	5,25	2,21
Median	5,00	2,25
Standard deviation	1,06	1,12
Standard error of mean	0.30	0.32

Table 12: Estimated number of warnings

B Test procedure script

B.1 Introduktion

Du kommer att köra ett scenarie i bilsimulatorn och därefter få fylla i ett par frågeformulär. Det hela beräknas ta 30-45 minuter.

Studien kommer inledas med att du får fylla i ett formulär med bakgrundsfrågor om körerfarenhet och liknande.

Du kommer sedan att få en introduktion till hur simulatorn fungerar samt prova att köra i den så du känner att du har kontroll över hur bilen beter sig.

Slutligen får du två frågeformulär på datorn att svara på.

Tänk på att du när som helst kan avbryta experimentet samt att alla uppgifter du lämnar kommer behandlas helt anonymt.

B.2 Instruktion Bakgrundsformulär

Börja med att fylla i formuläret under fliken ”Bakgrund”. Du fyller i fälten genom att dubbelklicka i dem och skriva. Om du vill göra en radbrytning måste du trycka in ”Alt” samtidigt som ”Enter” för att den inte ska byta fält.

B.3 Introduktion till simulatorn

Du kommer nu få köra ett scenarie utan trafik för att känna på hur simulatorn beter sig och svarar på kommando. Bilen är automatväxlad och ligger alltid i körläge, det finns alltså ingen back eller koppling utan bara gas och broms. Gaspedalen är till höger och bromspedalen är sammansatt av de två vänstra pedalerna för ökad stabilitet. Det kommer inte förekomma några avfarter under någon del av experimentet utan består helt enkelt av en landsväg med hastighetsbegränsning på 90 km/h där omkörning är förbjuden.

För att starta scenariot, tryck in den vänstra ”paddeln” bakom ratten och börja kör, tänk på att hastighetsbegränsningen är 90km/h. Denna del är ca 3 minuter lång, fråga gärna om det är något om körningen du undrar.

Efteråt: Om du vill kan du sträcka på benen innan vi fortsätter med huvuddelen

B.4 Introduktion till experimentscenarie

Du kommer nu att köra det längre scenariet, detta tar ca 13 minuter. Tänk åter på att hastighetsbegränsningen är 90km/h och omkörning är förbjuden. Innan du börjar vill jag att du tittar på skärmen till höger om ratten; när du hör och ser följande – (trigga distraktion) – vill jag att du upprepar de siffror som visas på den högt och tydligt, försök vara så noggrann du kan och få med alla samtidigt som du håller hastigheten och följer trafikreglerna.

Tänk på att köra som vanligt.

Finns det något du undrar över?

För att starta scenariot, tryck in den vänstra ”paddeln” bakom ratten och börja köra, tänk på att hastighetsbegränsningen är 90km/h. Denna del är ca 13 minuter lång.

Efteråt: Då vill jag bara att du svarar på ett par formulär så är vi klara.

B.5 Introduktion till NASA-TLX

Formuläret innehåller sex frågor. Läs varje fråga och markera på den tillhörande graderade skalan. Det finns inga svar som är rätt eller fel. Fundera inte för mycket på något påstående utan svara så som du tycker bäst passar in på hur du upplevde uppgiften. Skriv ett "x" i det fält om bäst stämmer överens med din känsla.

Säg till när du är klar med detta så ska jag presentera det sista formuläret.

B.6 Introduktion till trust-formulär

Nedanför finns en lista med påståenden rörande tillit till varningssystem. Varje fråga har en skala där du kan fylla i graden av din känsla av tillit till eller intryck av varningssystemet under körning. Vänligen skriv ett “x” i det fält som bäst stämmer överens med hur väl påståendet stämmer in med din uppfattning, det finns inget svar som är rätt eller fel.

C Background questionnaire

	Male	Female
Sex:		

Age:	

	0-2 år	2-5 år	5-10 år	10-20 år	>20 år
Hur länge har du haft körkort?					

	Dagligen	Varje vecka	Varje månad	Varje år	Aldrig
Hur ofta kör du bil?					

	Försiktig	Medel	Tuff
Hur skulle du beskriva din körstil?			

Vilken typ av resor gör du oftast med bilen?	

	Ja	Nej
Tror du att du skulle ha nytta av ett kollisionsvarningssystem?		

	Dagligen	Varje vecka	Varje månad	Varje år	Aldrig
Hur ofta tror du att du skulle behöva det per år?					

I vilken/vilka situationer tror du att du har behov? Beskriv	

	Ja	Nej
Har du kört någon bil som haft varningssystem installerat?		

Om ja, vilken typ av varningsystem?	
Om ja, har du upplevt några varningar?	

Table 13: Background questionnaire

D NASA-TLX and Trust questionnaires

Nedan följer sex frågor. Läs varje fråga och markera på den tillhörande graderade skalan. Det finns inga svar som är rätt eller fel. Fundera inte för mycket på något påstående utan svara så som du tycker bäst passar in på hur du upplevde uppgiften. Skriv ett "x" i det fält om bäst stämmer överens med din känsla

	Inte alls																			Väldigt mycket
Mental påfrestning: Hur mentalt påfrestande var uppgiften?																				
Fysisk påfrestning: Hur fysiskt påfrestande var uppgiften?																				
Tidsrelaterad påfrestning: Hur upplevde du tidspressen i uppgiften?																				
Prestation: Hur väl lyckades du utföra den genomförda uppgiften?																				
Ansträngning: Hur mycket behövde du anstränga dig för att klara uppgiften?																				
Frustration: Hur osäker, nedslagen, irriterad och stressad kände du dig?																				

Table 14: NASA-TLX Questionnaire

Nedanför finns en lista med påståenden rörande tillit och automatisering. Varje fråga har en skala där du kan fylla i graden av din känsla av tillit till eller intryck av varningssystemet under körning. Vänligen skriv ett “x” i det fält som bäst stämmer överens med hur väl påståendet stämmer in med din uppfattning.

	Stämmer inte Alls				Stämmer Helt
Grounds:					
1. Systemets sätt att fungera orsakar ett farligt eller skadligt resultat					
2. Systemet är konsistent					
3. Systemet är bedrägligt					
4. Systemet beter sig på ett vilseledande sätt					
5. Systemet ger ökad säkerhet					
6. Systemet varnar för ofta					
7. Systemet varnar för sällan					
Warrant:					
8. Jag saknade varningar i vissa situationer					
9. Jag tycker systemet gav för många varningar					
10. Jag förstår hur systemet fungerar					
11. Jag kan lita på systemet					
Backing:					
12. Jag är misstänksam till systemets avsikt, sätt att fungera eller utkomst					
13. Jag är känner mig van vid systemet					
14. Jag vill ha senare varningar					
15. Jag vill ha tidigare varningar					
16. Jag tror på systemet					
17. Jag känner mig osäker på systemet					
Rebuttals:					
18. Systemet fungerar bättre i vissa situationer					
19. Systemet fungerar sämre i vissa situationer					
20. Systemet störde min körning					
Questions related to system functionaity and experience:					
21. Jag tycker varningsinterfacet var intuitivt					
22. Jag tycker varningsinterfacet var oförståeligt					
23. Jag skulle rekommendera systemet till andra					
24. Systemet uppfyller mina förväntningar					
25. Hur många varningar skulle du uppskatta att du fick?					
26. Fanns det någon situation där du inte förstod varför det varnade? Om ja, hur många?					

Table 15: Trust questionnaire