Social and Emotional Characteristics of Speech-based In-Vehicle Information Systems: Impact on Attitude and Driving Behaviour

Ing-Marie Jonsson
At the Faculty of Arts and Science at Linköping University, research and doctoral studies are carried out within broad problem areas. Research is organized in interdisciplinary research environments and doctoral studies mainly in graduate schools. Jointly, they publish the series Linköping Studies in Arts and Science. This thesis comes from the Graduate School of Cognitive Science, a Division of Human Centered Systems at the Department of Computer and Information Science.

Distributed by:
Department of Computer and Information Science
Linköping University
581 83 Linköping

Ing-Marie Jonsson
Social and Emotional Characteristics of Speech-based In-Vehicle Information Systems: Impact on Attitude and Driving Behaviour

Upplaga 1:1
ISSN 0282-9800

©Ing-Marie Jonsson
Department of Computer and Information Science 2009

Tryckeri: LiU-Tryck
Abstract

Advances in modern microelectronics enable manufacturers to use advanced information systems in vehicles to provide and control a wide variety of functions and features. Even modest vehicles today are equipped with computer systems that control diverse functions from air-conditioning to high quality audio/video systems.

Since the primary task of driving involves the constant use of eyes and limbs, voice interaction has become an obvious means to communicate with in-vehicle computer systems both for control and to receive information. Perhaps because of the technical complexity involved in voice recognition, significant focus has been given to the issues of understanding a driver’s spoken commands. Comparatively the technology for voice reproduction is simple, but what effect does the choice of voice and its behaviour have on the driver? We know from human-human interaction that the timing and the social cues of the voice itself significantly influence attitude and interpretation of information.

Introducing speech based communication with the car changes the relationship between driver and vehicle. So quite simply, for in-vehicle information systems, does the spoken voice matter?

The work presented in this thesis studies the effects of the spoken voice in cars when used by in-vehicle information systems. A series of four experimental studies were used to answer the following questions: Do the characteristics of voices used by an in-vehicle system affect driver’s attitude? Do the characteristics of voice used by an in-vehicle system affect driver’s performance? Are social reactions to voice communication the same in the car environment as in the office environment?

The first two studies focused on driver emotion and properties of voices. The results show that the properties of voice interact with the emotional state of the driver and affect both attitude and driving performance. The third experiment studied the effect of voice on information accuracy. The results show that drivers’ perceptions of accuracy are dependent on the voice presenting the information and that this affects attitude as well as driving performance. The fourth study compared young and old drivers’ preferences for age of voice used by car information systems. Contrary to similarity attraction, the young voice was preferred by all drivers and had a positive influence on driving performance.

Taken together the studies presented in this thesis, show that both attitude and performance can be improved by selecting an appropriate voice. Results from these studies do not paint a complete picture, but they highlight the effects and importance of a number of voice related factors.

Results show that voices do matter! Voices trigger social and emotional effects that impact both attitude and driving performance. Moreover, there is not one effective voice or effective way of expressing information that works for all drivers. Therefore an in-vehicle system that knows its driver and possibly adapts to its driver can be the most effective. Finally, an interesting observation from these studies is that the social reactions to voice communication in the car are different than social reactions in the office. The so-called similarity attraction effects, an otherwise solid finding in social science, were not always found in these studies. It is hypothesized that this difference can be related to the different kinds of task demands when driving a car or working in an office environment.
Acknowledgements

It is hard to include in one page all the people that over the years have affected and influenced me over the years. I am deeply grateful to all of you, friends and critics, who have led me to this point in my life.

There are, however, a few people who deserve special mention. First and foremost is my principal advisor Professor Nils Dahlbäck, whose valuable guidance, patience and friendship has helped bring my work to fruition. I would also like to thank my other advisor, Assistant Professor Johan Åberg for his support, and Assistant Professor Olle Eriksson for help with methods and statistics.

I am indebted to my colleagues at Ericsson and to the Wallenberg Foundation that enabled me to move to Stanford University where the work on my thesis started. At Stanford I worked with Professor Clifford Nass and Professor Byron Reeves on social responses to communication technology. We were approached by Toyota InfoTechnology Center to explore new technology in vehicles. This led to a fruitful collaboration between Stanford University and Toyota InfoTechnology Center. I am especially grateful to Clifford Nass at Stanford University and Jack Endo (Endo-san) at Toyota InfoTechnology Center for support and help with advice, facilities and equipment to investigate speech systems using driving simulators.

Special thanks go to Dr. Mary Zajicek at Oxford Brookes University in Oxford, UK and Associate Professor Fang Chen at Chalmers Technical University in Gothenburg, Sweden with whom I have worked with over the years.

Among my friends I would like to give special thanks to Will and Mary Van Leer, Dr. Elizabeth Seamens, Dr. David Ofelt, Dr. Bjarne Däcker and Cristina Olsson. They have all helped make sure I never lost focus and stayed on the path.

This thesis is dedicated to my family; mother and father, Mariette and Tell Jonsson, who have provided so much generous love and support throughout my life, my brothers, Karl-Olof and Jan-Erik Jonsson, who have always been there for me, and of course, and most of all, to Ashley Saulsbury for the love, support, and kicking required to get me to finish what I promised so many years ago.
Contents

1 Introduction ............................................................................................................ 1
  1.1 Voices and Speech-based In-Vehicle Systems................................................ 1
  1.2 Background ..................................................................................................... 1
  1.3 Research Questions ....................................................................................... 17
  1.4 Methods used in Studies ................................................................................ 19
  1.5 Overview of Chapters .................................................................................... 28
2 Driver Emotion and Properties of Voices ............................................................ 29
  2.1 Emotions and Performance ........................................................................... 29
  2.2 Angry and Frustrated Drivers and Familiarity of Voice ............................... 34
  2.3 Matching Driver Emotion with Emotion of Voice ........................................ 53
  2.4 Making Driving and In-Vehicle Interfaces Safer and Better ....................... 65
  2.5 Stabilizing the Driver .................................................................................... 69
  2.6 Acknowledgements and Publications ............................................................ 71
3 Accuracy of Information in Vehicles................................................................... 73
  3.1 Driving, Trust and Quality of Information .................................................... 73
  3.2 In-vehicle Information and Hazard Warning System ................................... 75
  3.3 Design of Study ............................................................................................. 79
  3.4 Measures ........................................................................................................ 81
  3.5 Results ........................................................................................................... 83
  3.6 Discussion ..................................................................................................... 93
  3.7 Conclusion and Further Questions ................................................................. 97
  3.8 Acknowledgements and Publications .......................................................... 100
4 Older Adult Drivers and Voices for In-Vehicle Systems .................................. 101
  4.1 Older adults and Driving ............................................................................. 101
  4.2 Programs to Support Older Adult Drivers................................................... 103
  4.3 In-Vehicle Hazard and Warning System for Older Adults ......................... 105
  4.4 Assessing Two Voices for In-Vehicle System ............................................ 110
  4.5 Assessing the In-Vehicle Hazard and Warning System .............................. 116
  4.6 Conclusions ................................................................................................. 129
  4.7 Acknowledgements and Publications .......................................................... 132
5 Summary, Discussion and Conclusions ................................................................ 135
  5.1 Summary of Results from Previous Chapters ............................................. 135
  5.2 Summary of Additional Studies .................................................................. 137
  5.3 Emerging Patterns and General Observations ............................................ 146
  5.4 Limitations and Methods ............................................................................. 149
  5.5 Summary and Final Comments ................................................................... 158
6 References .......................................................................................................... 160
7 Appendices ......................................................................................................... 176
  7.1 Appendix A ................................................................................................. 176
  7.2 Appendix B ................................................................................................. 183
  7.3 Appendix C ................................................................................................. 189
1 Introduction

1.1 Voices and Speech-based In-Vehicle Systems

Automobile manufacturers, electronics and telecommunications companies are making computer based information systems available in all vehicles. Most cars today are fitted with interactive information systems including high quality audio/video systems, satellite navigation systems, hands-free telephony, and control over climate and car behaviour (Floudas, Amditis et al. 2004).

Even though most in-vehicle systems are screen-based, speech interactions are becoming more commonly used by in-vehicle systems. The use of speech technology in a vehicle would help increase the number of features and systems that can be controlled. There is limited space on steering wheel and dashboard for buttons. It would also enable drivers to keep their hands on the steering wheel and their eyes on the road during interactions with the system.

Speech communication with the car would also make the relationship between driver and vehicle very different from today. The social implications of introducing interactive media into the vehicle need to be studied. The aim of the work presented here is to study these effects in cars. Both as a general question of how results and findings from an office environment are applicable in driving environment, and also as targeted questions of how characteristics of voices such as gender, age, emotion and personality affect drivers’ attitude and driving behaviour.

More specifically, I address the following research questions in this thesis:

- Do voices matter?
  - Will characteristics of voices used by an in-vehicle system affect drivers’ attitude?
  - Will characteristics of voice used by an in-vehicle system affect drivers’ performance?
  - Are social reactions to voice communication the same in the car environment as in the office environment?

1.2 Background

In this section related work and background for two topics relevant to the rest of this thesis is discussed; the implementation of in-vehicle information systems and, how voice attributes influence listeners. With regard to the background information on in-vehicle systems the focus is on those employing speech-based interfaces rather than
the broader field of all in-vehicle computational systems. There is extensive information on how properties of speech and voices influence listeners. The related work on how characteristics of voices such as age, gender, personality and emotion influence attitude and performance is gathered mostly from psychology and media studies. However, the contexts for these studies are typically office and home environments. Furthermore, this section also describes previous work on how voices can be used to influence the perception of messages. Once again, the settings for these earlier studies were either office or home environments.

The background and related information presented in this section serves to highlight how properties of speech and voices have been found to influence listeners in contexts other than the driving environment. However, it also serves to introduce the questions central to this thesis; can properties of speech and voices be used to attract drives attention, and to focus and engage drivers by interactions with an in-vehicle system?

1.2.1 In-Vehicle Systems

Vehicles are often equipped with in-vehicle systems, either installed by the automobile manufacturers straight from the factory or, as after-market solutions by electronics and telecommunications companies. These systems include everything from high quality audio/video systems and satellite navigation systems to hands-free telephony and control over climate and car behaviour (Floudas, Amditis et al. 2004). Even though most in-vehicle systems today provide static road and traffic information based on maps, there are efforts to update the transportation infrastructure to increase driving safety and to give drivers more useful and timely information, such as road conditions, traffic situations and services. This type of intelligent-transportation system infrastructure will most likely provide connections and communications between vehicles and the roadside environment. Furthermore, intelligent systems such as Active Drive Assistance System (ADAS) (Bishop 2005), will be installed in vehicles. These systems are designed to help the driver to drive safely by providing traffic information, evaluate driver performance, and warn the driver of potentially dangerous situations. In some cases the system also takes control of the vehicle or part of the vehicle (ABS, tensing seatbelts, braking, deploying airbags).

In addition to safety and navigation systems, there is a focus on providing so called infotainment systems. These systems offer access to the vehicles conventional media systems (CD, radio etc), as well as new features such as Internet connections, including email and web browsing (Lee, Caven et al. 2001; Barón 2006). Many new services are initially provided by nomadic (portable) devices, such as mobile phones, navigation systems, Personal Digital Assistants (PDA), and MP3 players. For driver safety it becomes important to integrate these devices with existing in-vehicle
information systems and ADAS. In a recent European Union project, AIDE, the architecture of such an integration model was proposed (AIDE 2004-2008). This architecture will enable the driver to control all functions in the car using one interface, and for all devices to work together to provide the driver with the right information at the right time. Speech recognition is proposed as part of this architecture, and can be applied to in-vehicle functions in various ways.

1.2.2 Speech for In-Vehicle Systems

Automatic speech recognition technology can be used to input information and synthetic speech technology can be used for information output. Put together and used by a dialogue system these technologies would enable voice communication between driver and vehicle. The use of speech technology in the vehicle would solve two problems.

First; with the increasing number of features and systems that need to be controlled, there are a growing number of buttons and menus to be attended. Speech would help solve the screen real-estate problem since there is limited space on the steering wheel and dashboard.

Second; speech would enable drivers to keep their hands on the steering wheel and their eyes – and attention – on the road during interactions with the system. This is important since driving is the primary task when controlling a vehicle and driver distraction is generally defined as when a driver is performing a secondary task.

The single most important aspect of any system to be used in a vehicle is its impact on driving safety. Designers of in-vehicle information systems and devices must ensure that driver safety is preserved, and that drivers can keep their eyes and minds on the road and their hands on the wheel. There exists data to indicate that secondary task interactions while driving always lead to driver distraction and decreased driving performance (Barón 2006). Speech interfaces attempt to reduce physical distractions. However, even though speech interactions show some advantages over screen based interactions, these interactions will demand the driver’s attention and even simple conversation can disrupt attentive scanning and representation of a traffic scene. This is especially true in complex traffic situations or when driving conditions are bad.

1.2.3 Speech and Voice Characteristics

Sounds and speech can be used to direct a driver’s attention (Gross 1998; Gross 1999; Bower 2000; Clore, Wyer et al. 2001). Warning signals from the car can focus the driver’s attention to the dashboard, an utterance and a pointing finger by a passenger will direct attention to some object, and a honking horn will make the driver turn the
head to the sound. The amplitude, length or number of repetitions of a sound or a signal from a car or driving environment can be used to emphasize importance and urgency. The same effect could potentially be invoked by using different voices and changing the tone of voice in a speech based in-vehicle system.

Using verbal messages to inform or warn drivers could potentially be advantageous. Given that the language of the message is understood, a verbal message can be used to give the recipient more information than a signal (McIntyre and Nelson 1989). It can, for example, direct attention to different locations and suggest actions where a simple signal just indicates a fault. The potential danger here is that the system might trigger a driver reaction that would cause an accident. For instance, with a warning, the driver might step on the brake, and stop in an intersection creating a hazard for other motorists (Shahmehri, Chisalita et al. 2004).

When introducing computer generated speech messages in the car, it is also vital to address issues of “blind trust”. A consistent theme in today’s culture is that computers and interfaces cannot lie. Public perception is that they simply respond to the user’s performance consistently and objectively; they tell the user exactly what’s going on. This blind trust can itself lead to problems. There are reported incidents of drivers ignoring signs for road work and road closures, train tracks, and lakes, when following directions from navigation systems. In one case it became so bad that signs stating “Do Not Follow SAT NAV” went up in a village in the UK. How to present information to keep a driver’s trust and at the same time reduce incidents of “blind trust” becomes an issue for phrasing of information and selection of voices.

Choice of voice has long been an important factor for media companies that select TV and radio personalities. Results from media studies show that people unconsciously attribute human characteristics to communicating media and apply social rules and expectations accordingly. Using speech for in-vehicle systems highlights the potential influence of linguistic and paralinguistic cues. These cues play a critical role in determining human—human interactions where people respond to characteristics of voices as if they manifest emotions, personality, gender, and accents (Nass and Gong 2000; Tusing and Dillard 2000). An upset and loud voice can for instance be used to focus attention to a potentially dangerous situation. A happy and cheerful voice can potentially be used to put the driver in a better mood – happy people perform better than dissatisfied people (Isen, Daubman et al. 1987; Isen, Rosenzweig et al. 1991; Hirt, Melton et al. 1996; Isen 2000). A well-known and trustworthy voice may be used to convey important information – the benefits of trust include better task performance and willingness to use the system (Muir 1987; Lee and Moray 1994; Muir 1994)
People are often classified by how they speak and express themselves. Subsequent interactions are then affected by the interpretation of paralinguistic cues such as the rising tone of a question, the staccato of anger, or the familiarity of a voice you know. Cues can indicate affiliation and people are in general extremely skilful in determining others’ similarity to themselves after a few utterances. Homophily and similarity theories predict that people like voices that are similar to them (Byrne, Griffit et al. 1967; Byrne, Clore et al. 1986). This similarity is based on congruence of certain attributes in voice cues and choice of language, such as demographic variables, beliefs, values status, age, gender, class, education and occupation. Age is an important factor signalling affiliation. The interest in age cues for the driving environment is based on evidence that two groups of drivers are overrepresented in accident statistics. Drivers over 55 and drivers in the age group of 16-25 are involved in more incidents than drivers between 25 and 55 (these two groups are listed as groups at risk together with child passengers by the Center for Disease Control and Prevention (CDC), a US government agency). Finding cues or other properties of in-vehicle system to direct attention and support drivers in these age-groups would be desirable.

People are good at correctly determining the gender of a speaker. There are findings from social science that indicate a gender bias, so that female listeners prefer female voices and male listeners prefer male voices (Nass and Brave 2005). This should be balanced with findings from the aviation industry stating that female voices carry better in noisy environments (Nixon, Anderson et al. 1998; Nixon, Morris et al. 1998).

Emotions or moods are also associated with the acoustic properties of a voice (Cowie, Douglas-Cowie et al. 2001). Emotions influence peoples’ wellbeing, performance and judgment and can also divert or direct attention. Attention, performance, and judgment are important when driving, and even small disturbances can have enormous consequences. Considering that positive affect leads to better performance and less risk-taking—it is not surprising that research and experience demonstrate that happy drivers are better drivers (Groeger 2000). Emotional arousal is easy to detect in vocal communication, but voice also provides indications of valence through acoustic properties such as pitch range, rhythm, and amplitude or duration changes (Scherer 1989; Ball and Breese 2000). A bored or sad person will speak slower in a low-pitched voice, a happy person will exhibit fast and louder speech (Murray and Arnott 1993; Brave and Nass 2002), while a person experiencing fear or anger will speak with explicit enunciation (Picard 1997). Pre-recorded utterances, even though inflexible, are easily infused with affective tone. Cahn (Cahn 1990) has synthesized affective speech using a text-to-speech (TTS) system annotated with content-sensitive rules with acoustic qualities (including pitch, timing, and voice quality (Lai...
Introduction

2001; Nass, Foehr et al. 2001)). People were able to distinguish between six different emotions with about 50% accuracy, and people are 60% accurate in recognizing affect in human speech (Scherer, 1981). Affective state can also be indicated verbally through word and topic choice, as well as explicit statements of affect (e.g., “I’m happy), or with a sound. For example, fear is a reaction to a threatening situation, this could be a loud noise or a sudden movement towards the individual that results in a strong negative affective state, or preparation for actions to fight or flight. In an in-vehicle information system, unexpected sounds, such as a beep instead of “your tire pressure is low”, can activate a similar primitive emotional response. This mirrors how humans react to sounds that are disturbing or pleasing, such as screaming, crying, or laughing (Eisenberger, Lieberman et al. 2003). Emotional cues are furthermore an important set of cues since some emotions can be detected from voice in real-time (Jones and Jonsson 2005; Jones and Jonsson 2008).

People are extremely skilled at recognizing and tuning into a specific voice even when this voice is one of many - for example in a room full of people. Stevens (Stevens 2004) found that a particular brain region was involved in recognizing and discriminating voices. The right frontal parietal area is engaged in determining whether two voices were the same. Other studies found that familiar voices are processed differently than unfamiliar voices, and famous voices are recognized using different regions of the brain than when discriminating between unfamiliar voices (Van Lancker and Kreiman 1987; Van Lancker, Cummings et al. 1988; Van Lancker, Kreiman et al. 1989). Studies also show that the linguistic properties of speech (what is actually said) are processed in a different region of the brain than those regions that recognize and discriminate between voices (Kreiman and Van Lancker 1988; Glitsky, Polster et al. 1995). Together these studies show that voice discrimination is distinct and processed differently to what is actually said, even though conveyed in the same speech stream.

Familiar and famous voices are often used to emphasize or convince. Familiar is however also associated with loss of anonymity. Studies have shown a link between anonymity and aggressive driving (Ellison, Govern et al. 1995; Stuster 2004). The road-rage phenomenon (Joint 1995; Vest, Cohen et al. 1997; Ferguson 1998; James and Nahl 2000; Drews, Strayer et al. 2001; Fong, Frost et al. 2001; Galovski and Blanchard 2002; Wells-Parker, Ceminksy et al. 2002; Galovski and Blanchard 2004; Galovski, Malta et al. 2005) provides one undeniable example of the impact that emotion can have on the safety of the roadways.

Voice cues and choice of words can also signal personality. Cues such as loudness, fundamental frequency, frequency range, and speech-rate distinguish dominant from
submissive individuals have been shown to affect people interacting with systems using computer-generated speech (Manstetten, Krautter et al. 2001; Strayer and Johnston 2001). Even though cues for personality of speaker are less obvious and more subtle than cues for gender and age, people are generally very astute in interpreting the cues. Previous studies show that personality can be assessed by people using either linguistic cues and para-linguistic cues (Nass and Lee 2000). Literature shows for example that extroverts speak faster and with more pitch variation (paralinguistic cues) and also assertive language (linguistic cues).

1.2.4 Perception of Spoken Messages

Studies show that properties of speech affect how a message is processed and perceived. The primary characteristics that seem to cue these social responses are features of language such as personality (Nass and Brave 2005), interactivity (Nass and Moon 2000), and voice (Nass and Steuer 1993). Choice of words or phrasing of a message – linguistic cues - can also affect perception on messages. Linguistic cues can be seen as short signal phrases that indicate important information (Gaddy, van den Broek et al. 2001), and can hence be used to direct attention and affect interpretation, comprehension and attitude towards the message. There are a number of these cues that signal emotion or intention such as length of sentence (short for timid and longer for self-assured), repetition of utterances signal uncertainty and anxiety, and choice of words to signal everything from affiliation to attention and personality.

Female and male voices have the ability to influence the perception of a message in different ways (Tannen 1990; Nass, Moon et al. 1997). People tend to have gender based stereotypes where certain types of messages are better received using a female voice, and other messages are better presented using a male voice (Nass, Moon et al. 1997; Lee, Nass et al. 2000; Whipple and McManamon 2002). A study that tested listeners’ attitude towards different products when presented by a female or male voice found that the gender of the presenter’s voice does not affect gender neutral or male gender products, but has a strong effect on female gender products (Whipple and McManamon 2002). Their results show that a female voice worked better when the intended buyer was female, and that a male voice worked better when the intended buyer was male. Female voices are better at conveying emotional and caring messages and male voice are better at conveying instructional and technical messages (Nass, Moon et al. 1997).

Reaction times to recognize/categorize words are slower when two voices are used than when all the words are spoken using one voice (Mullennix and Pisoni 1990). This study also found that increasing the number of voices further slowed down the
time it took to recognize/categorize the recorded words. Similarly, examining the effects of familiarity of voice on recall for spoken word lists showed that lists produced by multiple voices lead to decreased recall accuracy. Words spoken by the same voice were recognised more often than words spoken by different voices (Goldinger 1996). Follow-up studies show that the advantage of single and familiar voices also holds for sentences (Nygaard and Fisoni 1998).

Famous people, and especially media people, are often trained in how to use their voices and can be better at reading and recording scripts needed to convey a message. Both radio and TV presenters are selected in part for their voices and how they talk. Furthermore, matching a famous or familiar voice to the content of a message could increase credibility and recall of the message (Papler 1974; Misra and Beatty 1990). A study where using a famous voice was compared to an unknown voice in an advertising campaign confirm these results (Leung and Kee 1999). This leads to the hypothesis that it could be advantageous to use familiar and/or famous voices for in-vehicle systems.

Accents and accented voices also influence perception and attitude. Using a French accent when talking about French wine instead for instance a German accent might influence buyers positively. However, findings from studies by Dahlbäck et al. (Dahlbäck, Wang et al. 2007) show that people prefer having tourist information given in an accent similar to their own, not given in an accent suggesting familiarity with the destination. Furthermore results in general show that accented voices are less intelligible than native voices, and that accented voices are less efficient than native voices for comprehension and retention (Tsali, DeShields et al. 1991; Mayer, Sobko et al. 2003). Accented speech was also found to be less comprehensible and harder to process when mixed with noise (Lane 1963; Munro and Derwing 1995; Munro 1998), making the use of accented voices in the noisy environment of in-vehicle systems less attractive.

Two important aspects of how voices and information influence messages in communication are similarity-attraction and consistency-attraction. Similarity-attraction predicts that people will be more attracted to people matching themselves than to those who mismatch. It has been applied to interactions with friends, business colleagues, partners, and computing applications. Similarity-attraction is a robust finding in both human-human and human-computer interaction (Byrne, Griffit et al. 1967; Nass, Moon et al. 1995; Nass and Moon 2000). In human-computer interactions, the theory predicts that users will be more comfortable with computer-based personas that exhibit properties that are similar to their own. Attraction leads to a desire for interaction and increased attention in both human-human (McCroskey,
Hamilton et al. 1974) and human-computer interaction (Lee and Nass 2003; Dahlbäck, Wang et al. 2007). In the same way, consistency-attraction predicts that people will like and prefer those who behave consistently. People are particularly sensitive to discrepancies between contents of a message and non-verbal cues (Ekman and Friesen 1974). Traditional media companies (TV, Radio, Movies) have long worked on establishing consistency in all aspects of presentation (Thomas and Johnston 1981). The reduced cognitive load and increased belief in a message resulting from consistency may make people more willing to interact with such a system. Results by Lee and Nass (Lee and Nass 2003) confirm these findings also in human-computer interaction. The authors investigated the effect of personality cues in voices and message content and show that similarity-attraction and consistency attraction holds. People felt better and were more willing to communicate when they heard a computer voice manifesting a personality similar to their own and using words consistent with their personality.

Voice characteristics have furthermore been found to have greater importance if the listener is less interested and involved in the topic; whereas voice matters less if the message is interesting. When both the content, interesting or non-interesting, and the voice, high intensity and intonation versus low intensity and intonation, was varied results show that voice characteristics matter when the message initially is not interesting (Gelinas-Chebat and Chebat 2001). Engaging qualities of voice characteristics with intensity and varied intonation has the potential to grab the listener’s attention even for low-engagement messages (Goldinger 1996). Goldinger (Goldinger 1996) investigated how changes in voices interacted with the focus of the listener’s attention, and found that changes in voice characteristics do not matter when the listener is focused on the meaning of the message. Conversely, when the listener is listening in a shallow manner, changes in voice characteristics could have a positive or detrimental effect on attention and recall.

1.2.5 Social Responses to Communicating Technology

Communicating with the car – especially if speech is used – will change the relationship between driver and vehicle. The social implications of interactive media have been explored by Byron Reeves and Clifford Nass. In their book “The Media Equation” (Reeves and Nass 1996), Reeves and Nass regard communicating media such as computers and television as inanimate objects, and demonstrate that despite this, people tend to react to them as if they were real people. They claim that most people, regardless of education and background, are faced with a confusion of real life and mediated life. Their findings show that peoples’ attitudes and behaviours when interacting with computers follow the same pattern as evidenced in social science findings (Reeves and Nass 1996).
Reeves' and Nass' studies on social response to communication media take them across topics such as communicating media and manners, communicating media and personality, communicating media and emotion, communicating media and social roles, and communicating media and form. In one of their first studies, for instance, they show that politeness is expected when interacting with computers – people are polite to computer and expect the computer to be polite in turn. Test subjects for this study denied that they would ever be polite to a computer, leading to the conclusion that their responses in the test were automatic and based on existing protocols for politeness.

This was followed by a study where they show how distance and interpersonal distance interacts with memory and perception. Close distance, big faces and local addresses (this computer is located in this building versus this computer is located in Chicago) makes people take more notice, trust the computer more and in turn be more truthful to the computer. Flattery, and specifically flattery by a computer, is another area that the authors investigated. Results from studies on computers that flatter their users show that people thought that they performed better and that they liked the computer more than when the computer did not flatter them. When faced with a survival task, results show that users perceive computers to have personality based on how they present themselves in text or voice. Furthermore, people with the same personality as the one projected by the computer system, worked better with and liked that system better than a computer system with miss-matched personality. In study after study, Reeves and Nass continue to use study protocols from social science where results show that people have some reaction to other people or the environment.

The Media Equation (Reeves and Nass 1996) is an interesting theory that has survived test after test to validate. It challenges common beliefs that people can consciously deal with differentiating real from fiction, similar to cognitive dissonance (Festinger 1957). People intellectually know that televisions and computers are inanimate objects but their behaviour does not always match. The Media Equation and the cognitive dissonance theory complement each other since the media equation causes a dissonance with how people react to television and computers. According to Festinger there must be an attitude change to reduce the dissonance, and according to Reeves and Nass this change in behaviour will not happen since a) it takes effort and b) it reduces the impact of the media experience. People react in an almost programmed way to television and computers, and in The Media Equation - Reeves and Nass have taken these reactions, studied them, and concluded the following: We know better than to scream at a television or a computer, but it takes too much effort to think about that while we are viewing the show or interacting with the program.
The majority of the research that follows The Media Equation can be considered to fall into four categories, reflecting the kinds of psychological or sociological effects that are being explored. These categories or areas of research in human-computer interaction explored in The Media Equation are a) traits, b) social rules and norms, c) identity, and d) communication. Research that focus on human traits includes studies on social facilitation (Rickenberg and Reeves 2000), social presence (Lee and Nass 2003), attraction (Nass and Lee 2000; Gong and Lai 2001; Lee and Nass 2003) and exploring the similarity attraction hypothesis (Byrne, Griffit et al. 1967; Byrne, Clore et al. 1986). Research concentrating on social rules and norms has studied reciprocity (Fogg and Nass 1997; Takeuchi, Katagiri et al. 1998; Nass and Moon 2000) flattery (Fogg and Nass 1997; Johnson, Gardner et al. 2004), and praise and criticism (Nass, Steuer et al. 2004). Research focusing on identity incorporates studies on group formation and affiliation (Nass, Fogg et al. 1996), and stereotyping (Nass, Moon et al. 1997). Nass, Moon and Green (Nass, Moon et al. 1997) show that both male and female users will apply gender-based stereotypes to a computer based on the gender of the computer voice. Research in communication has included studies exploring balance theory (Nakanishi, Nakazawa et al. 2003) and emotion theory and active listening (Klein, Moon et al. 2002). Results from this research, show that people experiencing negative affect felt better when interacting with a computer that provided sincere non-judgmental feedback.

The findings from these studies show that peoples’ attitudes and behaviours when interacting with computers follow the same pattern as evidenced in social science findings. Results show typical scripted human responses to communicating computers that implement characteristics such as gender, personality, group association, ethnicity, specialist-generalist associations, distance, politeness and reciprocity. For people to actively and consciously see computers as social participants in communication, at least one of three factors must be involved according to Reeves and Nass: (1) they must believe that computers should be treated like humans, (2) they respond to some human "behind" the computer when they communicate, or (3) people give the experimental researchers what they want - social responses. Prior to Reeves and Nass (Reeves and Nass 1996) the standard explanation for social responses to communicating computers was anthropomorphic – factors 1 and 2 (Turkle 1984; Winograd and Flores 1987).

A more compelling explanation than any of those above for people’s tendency to treat computers in a social manner is mindlessness. Please note that the term mindlessness is not a derogatory term, it simply means “automatically without reflecting and thinking” - indicating that people apply social rules and expectations to communicating with computers in the same way they do to communicating with
Introduction

people. Individuals respond mindlessly to computers since they apply social characteristics from human-human interaction to human-computer interaction based on contextual cues (Langer 1992). Instead of actively making decisions based on all relevant features of the situation, people that respond mindlessly draw overly simplistic conclusions – someone is communicating with me so I will apply all social rules that apply in this situation (even if it is a computer that interacts with me) (Nass and Moon 2000). In some situations, people are likely to show a stronger social response to humans than to computers. The majority of research conducted comparing people’s reactions to humans or computers have found a difference in the degree of the social reaction shown by participants, but no difference in the kind of reaction.

A study by Johnson, Gardner and Wiles (Johnson, Gardner et al. 2004) found evidence suggesting a link between degree of experience with computers and social responses to computers. An informal survey of computer users of varying levels of experience revealed that most people expect that users with high levels of experience with computers are less likely to exhibit the tendency to treat computers socially. This belief is based on the argument that more experienced users, having spent more time using computers, are more likely to view the computer as a tool. They are more likely to be aware of the computer’s true status, i.e. that of a machine. This argument shares the assumption inherent in both the computer as proxy and anthropomorphism explanations of the media equation effect: individuals’ social responses to technology are consistent with their beliefs about the technology. However, the research conducted did not support this argument, that is, Johnson et al. (Johnson, Gardner et al. 2004) found that more experienced participants were more likely to exhibit social responses. Specifically, participants with high computer experience reacted to flattery from a computer in a manner congruent with peoples’ reactions to flattery from other humans; the same was not true for participants with low computer experience. High experience participants tended to believe that the computer spoke the truth, had a more positive experience as a result of flattery, and judged the computer’s performance more favourably. These findings, considered in light of the “mindlessness” explanation of the media equation, highlight the possibility that more experienced users are more likely to treat computers as though they were human because they are more likely to be in a mindless state when working at the computer.

1.2.6 Speech and Driving Safety

In addition to investigating how different voices and different ways of expressing information affect attitude, it is also crucial to investigate if and how these cues affect performance. One of the most critical issues in evaluating in-vehicle systems is the demand for the driver’s limited attention. The driver’s primary task is safe driving; any other activity performed while driving is regarded as a secondary task. Driver
distraction is generally defined to be when a driver is performing a secondary task (Young, Regan et al. 2003). These tasks can be almost anything that is physical, visual or cognitive. Drivers have been observed reading, eating, putting on makeup and interacting with unsuitable or poorly located information devices while driving. Most of these distractions can be categorized as one the following:

1) performing a secondary task by moving the hands from the steering wheel (Barón 2006),
2) shifting the focus from the road to some information device (Barón 2006),
3) cognitive load induced by a secondary task disrupts scanning and comprehending road situations (Lee, Caven et al. 2001; McCarley, Vais et al. 2001; Strayer and Johnston 2001; Strayer, Drews et al. 2003),
4) the secondary task is more compelling than driving, causing full secondary task focus (Jonsson 2008).

Designers of in-vehicle information systems and devices should ensure that driver safety is preserved while interacting with these systems. Drivers should be able to keep their eyes and minds on the road, their hands on the wheel and fully focus on the driving task with a well designed in-vehicle system.

Do speech-based in-vehicle systems allow drivers to better focus on driving than screen-based in-vehicle systems? Current commercial in-vehicle systems rely almost exclusively on screen based interaction, often button or touch screen based, sometimes with speech output augmenting the screen based information. There are a few systems designed with speech interactions, these systems however, often also implement a screen-based interaction alternative. This convention, to implement a screen-based alternative to speech interactions in cars, is due to the car being a different (and less controlled) environment than the office for the use of speech technology. The vehicle presents a challenging environment where many factors, such as noise and the fact that drivers are often distracted or stressed, imposes new requirements on speech technologies.

The main difference between speech-based interactions in vehicles and most other environments is that the driver has to focus first on traffic and then on the speech system (Dahlbäck and Jönsson 2007). The fact that the driver does not pay full attention to the speech system alters the requirements of the speech systems dialogue management. Drivers might at any time in a dialogue pause to concentrate on the driving task, and when the traffic situation allows it, the driver should be able to resume the dialogue. The design of the in-vehicle dialogue system needs to be modified to handle the specifics of being a secondary task system and cope with interrupted and resumed interaction, repetitions, restart of dialogues, misrecognitions,
misunderstandings, presence and interruptions from other in-vehicle systems and passengers.

There are research projects and commercial products that can provide deeper insight into the design of dialogue system for in-vehicle information systems. As an example, there is VICO (Virtual Intelligent CO-driver) (Geutner, Steffens et al. 2002), a European project developing an advanced in-vehicle dialogue system. This dialogue system supports natural language speech interaction and provides services such as navigation, route planning, hotel and restaurants reservation, tourist information, car manual consultation. The system can adapt itself to a wide range of dialogues, allowing the driver to address any task and sub-task in any order using any appropriate linguistic form of expression (Bernsen 2002). Part of the European Union-funded TALK project (TALK 2004-2006) also focused on the development of new technologies for an dynamic and adaptive multimodal and multilingual in-vehicle dialogue system (Lemon and Gruenstein 2004). This dialogue system controls an MP3-player and supports natural, mixed-initiative interaction, with particular emphasis on multimodal turn-planning and natural language generation. Another example is DICO (Larsson and Villing 2006), a Vinnova (VINNOVA - 2009) project, that is focused on dialogue management techniques to handle user distraction, integrated multimodality, and noisy speech signals. The goal is to solve common problems in integrated dialogue systems such as common interface, clarification questions and switching between tasks in mid-conversation. The DARPA (Defense Advanced Research Projects Agency) in the US is sponsoring CU-move (Hansen 2000), that develops algorithms and technology for robust access to information via spoken dialog systems in mobile and hands-free environments. This project includes activities ranging from intelligent microphone arrays, auditory and speech enhancement methods, environmental noise characterization, to speech recognizer model adaptation methods for changing acoustic conditions in the car.

1.2.6.1 Commercial Speech-Based In-Vehicle Systems

Most commercial in-vehicle information systems are command based. In these systems interactions follow a strict menu structure where the driver gets a list of choices, and has to navigate through the menu structure step by step. One such system – Linguatronic, the first generation of in-vehicle speech systems - was introduced in 1996 in the S-Class Mercedes-Benz (Heisterkamp 2001). This system provided support for multiple languages and implements functions such as number dialling, number storing, user defined telephone directory, name dialling, and for operation of comfort electronics such as radio, CD-player/changer, and air conditioning. Since then many more speech systems have been deployed as aftermarket solution or by automobile manufacturers such as Fiat, BMW, and Honda.
Fiat worked with Microsoft to develop Blue&Me, a speaker independent in-vehicle infotainment system. Blue&Me is a driver initiated system with a push-to-talk button placed on the steering wheel. The Blue&Me system integrates in-car communication, entertainment and information, and includes support for mobile phones, mp3-players and GPS (Global Positioning System). The input is unimodal; the driver gives a voice command (for instance make to a phone call or to listen to a song). The output is multimodal; the system gives visual feedback on a dashboard display and auditory feedback via the car speakers.

BMW’s speech-based system is also a speaker independent push-to-talk system to control the radio, the phone the navigation system, and part of the iDrive system. Drivers can store phone numbers and names, and the system will use text-to-speech to read SMS and e-mails.

Honda’s push-to-talk system uses IBM’s voice recognition technology ViaVoice to control their navigation system. Drivers can ask for directions to a specific location or address, or ask the system to find local points of interest. The system supports request of the form “find the nearest gas station,” or “find an Italian restaurant in Los Gatos”. The system also enables control of the vehicle's climate system and audio/DVD entertainment system.

Experience with these in-vehicle systems shows that, even though speech recognition technology is a challenging area in the best of settings and conditions, the in-vehicle environment adds more complications. The car and in-vehicle environment have a wide variety of noises and usage patterns that confuse speech recognizers (Schmidt and Haulick 2006). Speech recognition errors are greatly increased by the noise that originates both from inside and outside of the vehicle. Noise from the engine, air conditioner, wind, music, echoes, etc, makes the signal-to noise ratio of the speech signal relatively low (Schmidt and Haulick 2006) making it harder for the recognizer to differentiate between words. Changes in speech patterns and inflections, due to the driver’s workload, stress and emotional state further reduces the speech recognition accuracy. Separating the driver’s speech from background noise is complicated by passengers talking, babies crying, children screaming and by sounds from passenger activities such as movies and mobile games. In this dynamic and changing environment, it is hard to find reliable patterns that indicate a particular speaker, and placing the microphone close to the driver’s mouth (headset) is not generally an option (Cristoforetti 2003; Chien 2005). It then often falls to the driver to correct recognition errors. This is both irritating and requires mental resources. If synthesized speech is used by the system, since comprehension of a synthetic message requires
more mental effort than comprehension of a spoken message (Lai 2001), the task becomes even harder.

1.2.6.2 Speech Systems and Driver Attention

Regardless of whether a system uses screen-based interactions, speech-based interactions or a mix thereof, these interaction tasks affect the driver’s attitude and driving performance. Screen-based interaction requires the driver’s eyes and focus to move from the road to the screen (Lunenfeld 1989; Srinivasan 1997). Recarte and Nunes (2000) also showed that mental tasks requiring operations with images produce more pronounced and different alterations in the visual search behaviour than those corresponding to verbal tasks. That different modalities use different cognitive resources was shown by Brooks in the 1960s (Brooks 1967; Brooks 1968; cited in Sanford 1985) Following this, Wickens (Wickens 1984) suggests that using speech-based interactions are less distracting since speech and visuals use different resources for attention and processing and driving is primarily a visual task. As a consequence of this, drivers can probably better divide attention cross-modally between ear and eye than intra-modally between two visual tasks (Wickens 1984).

Literature indicates that even speech based interactions with an in-vehicle system demand the driver’s attention with potential negative effects by reducing the driver’s on-road attention and increasing cognitive load. McCarley et al. (2001) demonstrates that simple conversation can disrupt attentive scanning and representation of a traffic scene. Drivers tended to take risks during speech interactions and often failed to compensate for slower reaction times (Horswill 1999). Lee et al. (2001) show that an in-vehicle information system that provides access to email while driving is perceived as distracting. Baron and Green (Barón 2006) reviewed and summarized papers on the use of speech interfaces for tasks such as music selection, email processing, dialling, and destination entry while driving. Most papers they reviewed focused on identifying differences between the speech and manual input modality from the viewpoint of safety and driver distraction. They concluded that “People generally drove at least as well, if not better (less lane variation, speed was steadier), when using speech interfaces than visual graphical interfaces”. The data the reviewed also showed that using a speech interface was often worse than just driving. Speech interfaces led to less workload than graphical interfaces and reduced eyes-off-the-road times, all pro-safety findings. Task completion time was less with speech interfaces, but not always (as in the case of manual phone dialling). Missing from the literature were firm conclusions about how the speech/manual recommendation varies with driving workload, recognizer accuracy, and driver age (Barón 2006). Lee et al. (2001) studied the effect of using an in-vehicle e-mail device (with simulated 100 percent speech recognition accuracy) on driver braking performance in a driving simulator. Self-
paced use of the speech recognition system was found to affect braking response time with a 30 percent increase in the time it took drivers to react to an intermittently braking lead vehicle. This demonstrated that speech-based interaction with an in-vehicle device increases the cognitive load on the driver.

Interactions with people show similar results, at least when the conversational partner is not in the car. Mobile phone conversations while driving show some of the same effects on driving performance. When using a mobile phone, part of the driver’s attention transfers from the road to the ongoing communication. This, together with the communication partner’s lack of knowledge of the driving conditions and the driver’s current situation, increases the risk of unintentionally creating a hazardous driving situation. Treffner’s study (Treffner and Barrett 2004) driving in real traffic confirmed that conversing on a mobile phone will detract from a driver’s ability to control a vehicle compared to driving in silence. It did not matter if the conversation was simple or complex or using a hands-free system, even speaking on a hands-free mobile phone while driving can still significantly degrade critical components of the perception–action cycle. These general result have been confirmed by numerous other studies investigating the impact of using mobile phones while driving (McKnight and McKnight 1993; Alm and Nilsson 2001; Strayer and Johnston 2001; Strayer, Drews et al. 2003; Kircher, Vogel et al. 2004; Strayer and Drews 2004). It is interesting to note that all these studies show increased response times to traffic events and that the use of hands-free phones does not strongly reduce distraction or response time (McKnight and McKnight 1993; Strayer and Johnston 2001; Strayer, Drews et al. 2003; Kircher, Vogel et al. 2004; Strayer and Drews 2004).

There are fundamental differences between listening to in-vehicle computers, conversations using mobile phones, and conversations with passengers. For passengers in the car, a study by Merat and Jamson (2005) show that there is a significant difference in the impact on a driver between a considerate and inconsiderate passenger. An inconsiderate passenger does not pay attention to the driver’s situation and workload and demands the driver’s attention during complex traffic situations. A considerate passenger, on the other hand, is sensitive to the driver’s workload and current driving conditions and traffic, and will refrain from interactions in situations where driver need to focus their full attention on the driving task.

1.3 Research Questions

From related work it is clear that introducing speech in the vehicle, even though speech-based systems have potential advantages over screen-based systems, will affect drivers’ behaviour. Care should be taken to design systems that are sensitive to
Introduction

the drivers’ situation and to design interactions that allow focus on the primary task – driving. Introducing speech-based in-vehicle information systems it is also important to address driver acceptance and usability in addition to driving safety. Especially since voices, speech and communication introduces social and attitudinal effects. Voices are not neutral! Voices carry a lot of socio-economic cues including indicators of gender, age, personality, emotional state, ethnicity, education and social status. The related work on voices and how they affect the attitude and perception emphasize the importance of these cues. Cues can potentially, when used appropriately, be used to direct attention, focus drivers, persuade drivers, and to build trust and liking. In the same way, cues can potentially, when selected inappropriately, annoy drivers, make drivers ignore messages, or focus drivers’ attention on (disliked) properties of the in-vehicle system instead of the intent of the messages.

Perception of information presented by the voice is influenced by the perception of the voice demographics, making it important to include the voice as a design parameter of in-vehicle systems. This is further complicated by the fact that different individuals perceive voices in different ways. A voice that is seen as positive by one individual can be perceived negatively by another. The negative impact of a voice is also potentially critical in an in-vehicle system since it can affect a driver’s performance as well as attitude. In the worst case, the effect on the driver could prove harmful for driving behaviour and driving safety, possibly even with a fatal outcome.

To investigate how voices and speech used by in-vehicle systems affects drivers we conducted a set of studies to address the following research questions:

- Do characteristics of voices such as age and emotional colouring used by an in-vehicle system affect drivers’ attitude?
- Do characteristics of voice used by an in-vehicle system affect drivers’ performance?
- Are social reactions to voice communication the same in the car environment as in the office environment?

There are large numbers of different in-vehicle systems and similarly, a large number of voice characteristics. The studies presented in this thesis do not aim to build a comprehensive map of drivers’ reactions to different voices. They are an effort to conduct explorative in-depth studies of selected in vehicles system and voice features to find out if voices matter and affect attitude and performance.
Below is a table with a non complete listing of different in-vehicle systems. From this table we selected to work with three types of systems: Navigation systems, Infotainment systems, and Hazard and Warning systems.

As can be seen from the table this involves two types of interaction models, purely informational system (the Hazard and Warning system) and interactive/Dialogue system (Navigation system, and Infotainment system).

Table 1-1: Types of In-Vehicle Systems

<table>
<thead>
<tr>
<th>Type of system</th>
<th>Interaction type</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Information</td>
</tr>
<tr>
<td>Navigation</td>
<td>x</td>
</tr>
<tr>
<td>ADAS/ Help/support</td>
<td>x</td>
</tr>
<tr>
<td>Infotainment</td>
<td>x</td>
</tr>
<tr>
<td>Hazard and Warning</td>
<td>x</td>
</tr>
</tbody>
</table>

For this thesis and in the studies reported in subsequent chapters, we selected a few voice characteristics to investigate. We have studied the effect of cues of affiliation and grouping based on gender of voice, age of voice, personality of voice, familiarity of voice and voice emotions. We also investigated the effect of accuracy of messages presented in a car. This particular property, accuracy, was selected based on the presumption that new information is selected and interpreted based on previous information from the same source.

1.4 Methods used in Studies

The driver’s primary task is safe driving. It is therefore crucial to investigate how speech-based in-vehicle information systems affect driving safety. It is also important to address driver acceptance and perceived usefulness of in-vehicle systems. What use is the best speech based in-vehicle system, if the driver does not like it and turns it off?

There is currently no standard mechanism to evaluate acceptance of new technology and new in-vehicle systems. Van Der Laan et al. (Van Der Laan 1997) proposed a tool for how to study the acceptance of new technology in vehicles. In their tool, driver experience is measured using a questionnaire with 9 items: useful/useless; pleasant/unpleasant; bad/good; nice/annoying; effective/superfluous; irritation/likeable; assisting/worthless; undesirable/desirable; and raising alertness/sleep-inducing. This tool can be used to rate the overall acceptance of a system, but there is no support to use the tool to diagnose and describe specific parts, such as a voice or a dialogue. There are published methods to evaluate interactive
speech systems (Graham 1999; Hone 2001; Larsen 2003; Dybkjær 2004; Zajicek 2005). There are as yet no standard methods that indicate how to measure the usability of an interactive speech-based in-vehicle system. It is desirable that methods are developed that take into consideration evaluating 1) the driver’s mental workload, 2) the distraction caused by the interactions with the system 3) how traffic interacts with the use of the system, 4) how passengers interact with the use of the system, and 5) drivers satisfaction and attitude. Guidelines as to which performance and attitudinal measures to use should also be published. The evaluation can take place either on a real road drive or using a driving simulator. To be able to compare results and measures used in different studies, certain testing conditions should be standardized, such as the participant screening and description, the fidelity level of the simulator, the traffic scenarios and the driving task (Jonsson 2006).

Common methods for driving performance measure the longitudinal acceleration or velocity, steering wheel behaviour or lane keeping displacement (Barón 2006). The driver’s visual behaviour during a driving session is normally measured using an eye tracking system to measure the eye glance pattern (Victor 2005; Barón 2006). To measure the driver’s mental workload the NASA-TLX method (Hart 1988) is normally used. The difficulty selecting driving performance measurement is that different drivers may use different behaviour strategies to cope with distractions. Some of them may reduce speed, others may position the car close to the right side of the road for a larger safety margin, and some may combine both behaviours. This can make data analysis difficult, and the results may not reflect the true situation. Interactions with an in-vehicle system can also be affected by changes in the driver’s mental workload due to exterior factors such as traffic or road conditions. During complex traffic situations, even simple speech tasks may significantly increase the mental workload and result in decreased driving performance. During light traffic and easy road conditions, the driver may be able to use more resources to cope with the in-vehicle system. An in-vehicle system can potentially also keep the driver alert resulting in improved driving performance, for instance engaging drowsy drivers in limited interactions. Different types of speech-based in-vehicle systems, such as light interaction, complex dialogues, or purely informational systems, may also impact driving performance differently. It might therefore be necessary to develop special methods, tailor-made to continually measures workload for drivers (Wilson 2002; Wilson 2002) in addition to the NASA-TLX. For driving safety reasons, new methods are best tested in a driving simulator, and then verified in real traffic.

1.4.1 Driving Simulators as Tools

All studies in this thesis are done using a driving simulator, and the results constitute an indication of behaviour in real cars and real traffic, but no guarantee.
There are many factors that influence the choice of a driving simulator for initial testing. Driving is a complex activity that continually tests drivers’ abilities to react to the actions of other drivers, traffic and weather conditions, not to mention unexpected obstacles. Despite the dangers involved in driving, the average driver will have very few accidents in their lifetime. While many of these incidents do not result in serious injury, some do cause harm and even death. Because of the rarity of accidents, it would be too time consuming to set-up an experiment with the characteristics of real driving and wait for a significant number of events to occur. On the other hand it would be impractical, given the liability for safety, to study driving behaviour by subjecting people to high-risk real-life driving. Therefore, the best way to examine accidents is to challenge people within a driving simulator. The experience is immersive, to different degrees depending on the fidelity of the simulator. The simulator can be programmed to subject drivers to more risky situations in 30 minutes of driving than they would be within a lifetime of driving. At the same time, people are spared the psychological and physical harm that comes with real accidents.

Two driving simulators were used in these experiments. A video game, a PlayStation2 running Hot Pursuit, was used for two studies. All other studies used a commercial driving simulator, STISIM Drive model 100 with a 45 degree driver field-of-view, from Systems Technology Inc. In all studies users sat in a real car seat and “drove” using a Microsoft Sidewinder steering wheel and pedals (consisting of accelerator and brake). The simulated journey was projected on a wall in front of participants.

Hot Pursuit was used for the first study (described in chapter 2). The video game was configured with pre-programmed settings for car, driving conditions and driving course. The screen was videotaped for later manual coding of driving behaviour.

Horn-honks were generated at preset intervals to measure attention to driving task. The number of responses and response times to these horn-honks were automatically recorded. All verbal utterances by the drivers were also recorded for later analysis.
Introduction

The simulator properties were set to be the same for all participants in a study. All drivers used the same car, thereby experiencing the same vehicle properties such as acceleration, brakes, and traction. All drivers drove in the same weather and daytime setting. And all drivers in a study completed the exact same driving scenario (same road layout and same driving environment down to the colour of cars and houses), even though they were assigned different conditions based on the properties of the in-vehicle information system.

Depicted to the left is a road with signs and traffic. Note the rear-view mirror located in the top right corner of the picture. Traffic (at the level of individual cars) can either be programmed to follow traffic regulations or drive without adherence to traffic regulations.

To the left is a screenshot from the simulator that show an intersection with traffic lights. Intersections can be defined as full intersections or T-intersections (left or right); they can have no signage, stop signs, yield signs or traffic lights.
There are some differences between driving scenarios in the Hot Pursuit setup and STISIM Drive. A driving scenario in Hot Pursuit is static and takes the driver around a predetermined track. The length of the driving session was set by the in-vehicle system. Drivers could therefore, depending on speed, complete a different number of laps around the track.

A driving course in STISIM Drive is described by defining a road and placing objects along that road. Roads are defined in terms of length, number of lanes, and vertical and horizontal curvature. Intersections, signage, houses, pedestrians and cars are placed along the road in locations specified by distance from the beginning of the driving course. Cars can be parked, driving in the same direction as the test-driver, driving in the opposing direction or intercepting the test-driver at intersections. A driving scenario in STISIM Drive is also static and predetermined, and can be programmed to have a specific length. Drivers can turn left or right at any intersection, but will still be driving on the same road as if they had continued straight ahead. This ensures that all drivers experience the same road regardless of turns and that all drivers take the exact same road once from start to finish.
Introduction

The in-vehicle system is programmed to interact at certain locations along the road. These features of STISIM Drive ensure task consistency. It guarantees that all participants drive the same route for the same distance and interact with the system at the same locations in the driving scenario.

The audio output from the in-vehicle information systems was played out of speakers in front of the driver, mimicking the sounds that would come from speakers on the dashboard. For each study, the amplitude of the in-vehicle system was set by pilot subjects, and then kept at the same level for all participants in that study. This resulted in noticeably louder settings in driving experiments with older adults than with age groups 18-25, since older people find it more difficult to distinguish speech in noisy environments (Gordon-Salent and Fitzgibbon 1999).

All participants in the studies started with a 5 minute test run of the simulator to familiarize themselves with the workings and the controls. This enabled participants to experience feedback from the steering wheel, the effects of the accelerator and brake pedals, a crash, and for us to screen for participants with simulator sickness (Bertin et al. 2004). The test run is particularly important for older adult drivers and previous studies show that older adults need about three minutes of driving to adapt to the simulator (McGehee et al., 2001).
Using a driving simulator is not an entirely realistic driving environment. The realism varies with the fidelity of the driving simulator, and some of the most advanced simulators reproduce a 360 degree visual and full tactile and kinaesthetic feedback. All studies presented here have been done using rather simple driving simulators. The experience of driving in a simulator differs from driving a real car; see the phenomena of driving simulator sickness. This raises the question of how valid results obtained in simulators are compared to real driving. I will return to this issue in chapter 5.

1.4.2 Driving Performance Measures

While it may seem logical to focus on the nuances of steering, lane position, acceleration and braking patterns, this information is not readily available for real drivers. In fact, while the number of accidents that drivers have and the number of moving violations may seem crude, these figures are exactly the numbers that private and public agencies use to evaluate drivers. The US Department of Motor Vehicles (DMV) uses a point system based on speeding and accidents to judge whether or not drivers should keep their licenses. Similarly, insurance companies use similar measures to determine premiums. While it would be valuable to have a thorough understanding of all nuances of driving behaviours and how they impact driving performance over longer distances and time, it is not practicable in real traffic or in simulated scenarios.

What is important here is that the crude numbers of driving accidents and violations serve the same function in both real and simulated driving. The numbers indicate critical breakdowns in driver attention, judgment, and vehicle management; it is these failings that predict future driving problems. Thus, when looking at behaviour in driving simulators, as well as in real traffic, we look to accidents, speeding, and traffic violations in addition to lane and brake behaviour to give us valuable insight into patterns of driving behaviour.

For most studies, I have therefore focused on measures for the most dangerous behaviours: number of collisions, number of off-road accidents, swerving, and obedience to the most important traffic laws (adherence to traffic lights and adherence to stop signs). As mentioned before, all of these negative behaviours are much more common in a driving simulator than in actual driving and one key reason is that we can (and do) create extremely difficult driving courses as a basis to study variance in driving performance: A simple driving course of the same length would fail to generate any variance in poor driving behaviour.
1.4.3 Attitudinal Measures

The approach to attitudinal measures and the selection of scales used in the studies presented here are influenced by The Media Equation (Reeves and Nass 1996). Based on the theory and my privilege to work with the authors, I believe that the social dynamics surrounding human-human interactions are shown to exist in human-computer interactions. The studies described in the upcoming chapters use attitudinal measures from communication research and psychology. This includes standard measures of blame attribution, emotional state, personality, homophily, willingness to communicate, source credibility and trust (Rubin, Palmgren et al. 1994).

While the studies presented in The Media Equation were all performed in an office setting and provided only attitudinal measurements, the hypothesis is that the same attitudinal responses would hold (outside the office setting) in the context of in-vehicle information systems. In the studies presented in this thesis I have added performance and behaviour measures in addition to the attitudinals. I demonstrate that attitudinal measures interact with performance and behaviour measures. Most importantly I study if the driving environment is indeed the same as the office setting and if results from the office setting presented in The Media Equation still hold for drivers and cars.

1.4.4 Inducement Techniques

Word choice and linguistic cues can be used to direct attention and affect interpretation, comprehension and attitude towards the message (Gaddy, van den Broek et al. 2001). A strategy used in many studies is “self-disclosure” (Jourard 1926-1974). This is a simple linguistic cue that can aid communication by sharing information about oneself, history, present, emotions and thoughts. Even though it is a simple approach, it has the potential to improve intimacy, rapport in face to face communication, and even improve public speaking and connecting with groups. When a system shares information about itself, it allows itself to be “seen”, and it makes it easier for drivers to relate to that system. Once a communication partner engages in self-disclosure, normal social behaviour leads to the other communication partner disclosing information. This is known as the norm of reciprocity. Mutual disclosure makes the communication more personal, deepens trust in relationships. Communication partners feel better about themselves and the interactions. When people perceive a system to be more human-like and not entirely depersonalized, communication, and relationships improve. This strategy was used in studies described in more detail in chapters 2 and 5 to entice drivers to interact with the system.
For studies where drivers were expected to be subject to particular emotions, there are two methods to achieve this a) work with the emotions that participants bring to the lab or b) create the emotion once they enter. While there are advantages and disadvantages to both approaches, the difficulty of scheduling equal numbers of emotional drivers as well as the desire to ensure that the emotion experienced by participants be created in the same manner led us to create emotion rather than rely on recruitment (Masters, Jonsson et al. 2006). Therefore, the target emotions were induced in each participant at the beginning of the experiment using a variety of inducement techniques such as video-clips, computer-tasks and visualizations techniques. These techniques are described in more detail in chapter 2.

1.4.5 Statistics

The main method for analysis used in this thesis is analysis of variance (ANOVA). There are several types of ANOVA depending on the number of treatments and the way they are applied to the subjects in the experiment. The most commonly used types of factorial ANOVA in this thesis are the 2x2 and 2x3 designs. Reported from these ANOVA’s are the main effects and interaction effects (Moore and McCabe 1998).

Linear regression analysis is used for one of the studies, where the independent variable lent itself to be seen as a continuous value instead of as discrete groups. For this particular study, the response to the independent variable assumed a linear function of behavioural and attitudinal behaviours (Pedhazur 1997).

When appropriate, paired samples t-tests is used to compare two variables for a single group, and independent-samples t-tests are used to compare means for two groups of cases (Moore and McCabe 1998).

Cronbach’s α is used to measure the reliability of an index when the index is developed from combining a set of items or questions. The measure indicates how well the items of the index are correlated, and is an indication that the index provides stable responses over repeated administrations (Cronbach 1951).

To assess the effect size of the measures, I use Partial $\eta^2$ which is one of the methods recommended for use in factorial designs to ascertain effect size for each component (Pedhazur 1997). It specifically measures the proportion of variance accounted for by each component. Components associated with high values of Partial $\eta^2$ account for a large proportion of the variance and are hence more useful (powerful) to explain a behaviour/phenomena than a component with a lower value of Partial $\eta^2$. 
1.5 Overview of Chapters

This thesis is based on a number of studies where the effects of variations in voices and linguistics of in-vehicle systems on driver’s attitude and behaviour have been studied. These studies are organized as follows: There are four additional chapters in this thesis, three of these chapters describe my studies of how characteristics of in-vehicle systems affect drivers in more detail, and the last chapter provides a summary of my work, including more studies, and discusses the results in a wider context. The three chapters with details on studies of in-vehicle systems are written so that they can be read separately.

Chapter 2 describes two studies with emotional drivers and how they are affected by characteristics of voices. Previous results show that emotional drivers have the potential to be good drivers (happy drivers) or bad drivers (road rage). It is important to investigate if in-vehicle systems can be used to improve driving performance by influencing drivers’ emotional state.

Chapter 3 gives a detailed description of the study on how accuracy of information affects drivers. This is an important aspect of in-vehicle systems since it affects trust and reliability.

Chapter 4 contains a detailed description of a two studies on selecting voices for in-vehicle systems for two groups of drivers, older adult drivers and young drivers under the age of 25. It is important to investigate the effect of in-vehicle systems on these two age groups since they are overrepresented in accident statistics. In particular, older adult drivers can either be substantially helped by a system that is well-designed or driven to distraction by a less appropriate system.

Chapter 5 contains a summary of the results from my studies on speech based in-vehicle systems not described in detail in this thesis. There is also a discussion on the validity and generalization of the results. Outlined are also more research questions and other areas of speech based in-vehicle systems to investigate.
2 Driver Emotion and Properties of Voices

2.1 Emotions and Performance

People are affected by traits and states as well as by the environment. State is defined to be the feelings, knowledge, and physical situation of the individual. Prototypical states are temporary, brief. Traits are defined to be age, gender and personality. Prototypical traits are stable, long-lasting, and internally caused. Traits and states are concepts that people use to both describe and understand themselves and others. In all our activities, states and traits influence everything we do and experience, from answering the phone to driving down the highway. To be able to predict an individual’s behaviour at any given moment in time requires attention to state as well as traits (Watson and Clark 1994).

In this chapter I present two studies where we took an in-depth look on how state and in particular emotions of drivers interact with characteristics of voice. The voice characteristics considered in these studies were emotions and familiarity.

2.1.1 Emotions

Emotion is a fundamental component of being human and motivates actions that add meaning to our experiences. Emotion is more complex than simple excitement when a hard task is resolved or frustration when reading an incomprehensible error message. Literature on emotion has grown during the past few years, and new results show that emotions play a critical role in all goal-directed activities (Heberlein and Adolphs 2004), including driving. There are five generally agreed upon aspects of emotion that stand out:

1. Basic emotions typically include fear, anger, joy, disgust, and sometimes also interest and surprise (Ekman 1992)
2. Emotions are a reaction to events deemed relevant to the needs, goals, or concerns of an individual (Kleinginna 1981)
3. Emotion encompasses physiological, affective, behavioural, and cognitive components. (Kleinginna 1981)
4. Emotions are essential characteristics possessed at birth and address specific environmental concerns and each emotion is associated with a specific set of physiological and cognitive responses (Tooby 1990).
5. All emotions except the primary/primitive emotions are learned social constructions (Tooby 1990; Shweder 1994) and emotions are likely to vary across cultures, and cross-cultural consistencies are based on a common social structures rather than biology.
In neuropsychology, a common model of the brain has three regions that are involved with emotions (LeDoux 1996). These areas are the thalamus, the cortex and the limbic system defined as the hypothalamus, the hippocampus, and the amygdala. The thalamus receives input from the environment that it then sends both to the cortex and to the limbic system. The limbic system evaluates the inputs relevance to the individual’s goals and needs, and if the input is considered relevant, the limbic system signals the body for physiological responses and the cortex for cognitive processing.

Some “primitive” or “primary” emotions, such as startle-based fear, aversions and attractions, are based on a link between the thalamus and the limbic system (Damasio 1994). For example, fear is a reaction to a situation that has the potential to threaten an individual. This could be a loud noise or a sudden movement towards the individual that results in a strong negative affective state, as well as physiological and cognitive preparation for actions to fight or flight. In the context of a voice based in-vehicle information system, unexpected sounds, such as a beep instead of “your tire pressure is low”, have the potential to activate a similar primitive emotional response. This mirrors how humans react to sounds that are disturbing or pleasing, such as screaming, crying, or laughing (Eisenberger, Lieberman et al. 2003).

An emotion can also result from a combination of both the thalamic-limbic and the cortico-limbic pathways. An event causing an initial fear reaction can be later recognized as harmless by more extensive, rational evaluation such as when you realize that your car’s sudden beep is just meant to catch your attention, to the fact that your tire pressure is low.

Individuals communicate most of their emotions by a combination of words, sounds, facial expressions, and gestures. Anger, for example, causes many people to frown and yell. People also learn ways of showing their emotions from social interactions, though some emotional behaviour might be innate. Paralinguistic cues such as tone of voice and prosodic style are among the most powerful of these social signals even though people are usually unaware of them. Off the shelf technologies can be used to assess, detect and identify emotions or emotional states in real-time (Picard 1997).

Differentiating between emotion and mood is important since they influence interactions differently. The difference between emotion and mood is time. Mood is a longer term state that biases people’s responses, whereas emotion is a more immediate and short duration affective state (Davidson 1994). Mood has a different impact on attention and people tend to pay more attention to events and visuals that are relevant to their current mood (Bower 2000). It has also been shown that people
that are in a good mood often regulate mood by performing tasks that sustain their mood or counteract undesired moods.

Emotions also impact memory; emotional events and visuals are better remembered than unemotional events and visuals. It is interesting to note that negative events are remembered better than positive events (Reeves and Nass 1996). Memory is also affected by mood, and follows the consistency theory; people in a good mood will remember happy events better than sad events. It is interesting to note that a positive mood decreases risk-taking. This might be in an effort to preserve the positive mood. So, even if people in a positive mood are more risk-prone when making hypothetical decisions, they tend to be more cautious when presented with an actual risk situation (Isen 2000).

People also often “catch” other’s emotions, such as when a person becomes happier when communicating with a person that is laughing and happy. Since we don’t have a set of social rules for interaction with computers yet (Reeves and Nass 1996) emotions in in-vehicle interfaces has the potential to be contagious in the same way.

2.1.2 Emotions and driving

Driving in particular presents a context in which emotion can have enormous consequences. Attention, performance, and judgment are of paramount importance in automobile operation, with even the smallest disturbance potentially having grave repercussions. The road-rage phenomenon (Galovski and Blanchard 2002; Galovski and Blanchard 2004; Galovski, Malta et al. 2005) provides one undeniable example of the impact that emotion can have on the safety of the roadways. Considering the above discussion of the effects of emotion—in particular, that positive affect leads to better performance and less risk-taking—it is not surprising that research and experience demonstrate that happy drivers are better drivers (Groeger 2000).

Now that car manufacturers are turning to voice as a strategy for interactions with everything from navigation systems and environmental controls to road-aware co-pilots, it is critical to know how a driver’s emotion interacts with characteristics of an in-vehicle voice interface in affecting attention, performance, and judgment.

2.1.2.1 Attention

Emotions can direct our attention to objects and situations that are important to our needs and goals. This is done through emotion-relevant thoughts that dominate our conscious processing. The focus increases with the importance of the situation (Clore and Gasper 2000). This attention-getting function can be used in a positive way by an in-vehicle information system to alert the driver by generating “turn left now”, or it
can be distracting when drivers are frustrated by poor voice recognition and can think about nothing else. People tend to pay more attention to thoughts and stimuli that have some relevance to their current emotion (Bower and Forgas 2000), so it is important for an in-vehicle system to focus attention and follow-up interactions in a positive way.

Just as emotions can direct users to aspects of an in-vehicle information system, emotions can also drive attention away from the stimulus eliciting the emotion (Gross 1998). For example, becoming angry with the voice recognition part of an in-vehicle information system may be seen as un-productive since the system cannot possibly realize that the driver is upset. An angered driver may subsequently try to avoid parts of the system that rely on voice input, rendering the driver’s interaction less efficient or effective. In extreme cases, the user will simply turn the system off. If the emotion is too strong, however, the driver will not be able to ignore the source (Wegner 1994), potentially even resulting in rage. Positive emotions may likewise require regulation at times, as when amusing content, such as a joke leads to laughter at an inappropriate time and place.

2.1.2.2 Performance
Emotion has also been found to influence performance. The most striking finding is that even mildly positive feelings profoundly affect the flexibility and efficiency of thinking and problem solving (Murray, Sujan et al. 1990; Hirt, Melton et al. 1996; Isen 2000). In one of the best-known experiments, subjects were induced into a good or bad mood and then asked to solve Duncker’s candle task (Duncker 1945). Participants were given only a box of thumbtacks and had to attach a lit candle to a wall such that no wax could drip on the floor. The solution requires the creative insight to thumbtack the box itself to the wall and then tack the candle to the box. Participants who were first put into a good mood were significantly more successful at solving this problem (Isen, Daubman et al. 1987). In another study, medical students were asked to diagnose patients based on X-rays after first being put into a positive, negative, or neutral mood. Participants in the positive-affect condition reached the correct conclusion faster than subjects in other conditions (Isen, Rosenzweig et al. 1991). Conversely, positive affect has been shown to increase reliance on stereotypes and other simplifying rules of processing, which could lead happy users to make less nuanced judgments about a voice interface and to be more influenced by labels, such as the gender, personality, and accent of the voice (Schwartz and Bless 1991; Reeves and Nass 1996; Isen 2000).
2.1.2.3 Judgment

Emotion can influence judgment and decision making, and, as mentioned above, emotion tends to bias attention and thoughts in an emotion-consistent direction. One important consequence of this is that everything—even those people, things, and events, that are unrelated to the current affective state—is judged through the filter of emotion (Niedenthal, Setterlund et al. 1994; Clore and Gasper 2000; Erber and Erber 2001). This suggests that drivers in a good mood would most likely judge both an in-vehicle system itself, as well as what the in-vehicle system says, more positively than if they were in a negative or neutral mood. Recommendations would hence obtain a greater level of acceptance among happy people. Positive emotion also decreases risk-taking so that even though people in a positive mood are more risk-prone when making hypothetical decisions, they tend to be more cautious (Isen 2000) when presented with an actual risk situation.

2.1.3 Inducing Emotions

When running experiments concerning emotion there are basically three options a) work with the emotion that participants bring into the lab, b) acting emotions, and c) create the emotion once they enter the laboratory. There are advantages and disadvantages with all of these approaches.

It is hard and time consuming to rely on participants bringing a certain emotion to the lab. This is imprecise and unpredictable and requires scheduling and re-scheduling if the emotion changed from scheduling to arrival at lab.

Literature indicates that induced or “natural” emotion is superior to acted because it represents how people actually behave (Masters, Jonsson et al. 2006). For applications where we must model genuine emotions, induced or observed emotions should therefore be used.

Induced or natural emotion data can be acquired in numerous ways. Subjects can be induced by participating in a situation where people are engaged in emotional discussions, either live with members of a research team or as seen on a TV program (Masters, Jonsson et al. 2006). Emotions can also be induced by having subjects do readings of emotional passages (7-8 sentences) (Masters, Jonsson et al. 2006) or by using video clips as stimuli. Detenber & Reeves (Detenber and Reeves 1996) suggest that still pictures and image video clips influences emotional state. This is the strategy we used to induce emotions in one of the studies reported in this chapter.

Depending on the desired emotion to induce however, video clips may not always be optimal, for instance Gross and Levenson (Gross and Levenson 1995) reported
difficulties eliciting high levels of reported anger. Moreover, films designed to elicit anger states often elicit a blend of negative emotions, including related states such as disgust and sadness. Films are here at a disadvantage relative to techniques that induce anger through interpersonal situations. The explanation for this is most likely that anger requires a high level of personal engagement and/or immediacy and this is hard to achieve with a film. Other methods to induce anger include computer based tasks, such as the Stroop colour-naming and math tasks (Stafford 2003), and visualization tasks (Fessler, Pillsworth et al. 2004).

For the studies presented in this chapter we wanted to ensure that the emotions experienced by participants were created in the same manner. This led us to induce or create emotion rather than rely on recruitment.

2.1.4 Driver Emotion and In-Vehicle Information Systems

The car presents an environment where speech based information systems can be both an aid for the primary task of driving (hazard and warning system, navigation etc.), and an information tool to enhance the driving experience (personalized information, recommendations for routes, restaurants, etc). Sounds or characteristics of an in-vehicle voice have the potential to impact the drivers focus, attention, performance, and judgment. The sound, voice, dialogue system and sentence structure used for in-vehicle information systems must therefore be designed with several considerations in mind. Should warning signals be used instead of voice prompts for critical systems? Should the voice of a familiar or famous person be used for credibility? Should a female or male voice be used matching drivers? Should an accented voice be used for content matched messages? Should an emotional voice be used to emphasize the message or to induce emotions desired for driving safety?

The studies presented in this chapter focus on two of these aspects:

- Should a familiar voice be used for angry and upset drivers, and if so what are the effects? Will the familiarity of the voice and its influence on anonymity, believability and trust, influence the driver’s behaviour and attitude?
- Should an emotional voice be used to match the emotional state of the driver, and if so what are the effects? Might the emotional characteristics of the voice have as much impact on attention, performance, and judgment as the emotion of the driver?

2.2 Angry and Frustrated Drivers and Familiarity of Voice

When the American Automobile Association (AAA) contracted the Gallup Organization to investigate driver concerns', they found that motorists felt more
threatened by aggressive drivers than by drunk drivers; 40% of the respondents said that aggressive drivers "most endanger traffic safety," while 33% identified drunk drivers as the primary risk (Joint 1995; Connell and Joint 1996; Mizell 1997). Most drivers have either been subjected to aggressive drivers or experienced road rage themselves.

A review of 10,037 aggressive driving incidents (Mizell 1997) from newspapers, police reports, and insurance reports clearly illustrates that there is no one profile of the so-called "aggressive driver." Although the majority of the perpetrators are between the ages of 18 and 26, there are hundreds of cases in which the perpetrator was 26 to 50 years old. In 86 known cases (1990 to 1996) the aggressive driver was 50 to 75 years old (Mizell 1997). There is also a clear gender difference in reported aggressive driving incidents, with only 413 confirmed female drivers (528 potential – gender unknown) of the 10,037 reviewed incidents (4%-9%).

Although aggressive behaviour can be sparked by trivial events -- "He stole my parking space", "She cut me off" – they are rarely the result of a single incident but rather are in reality the cumulative result of a series of stressors in the motorist's life. The traffic incident that turns violent is often "the straw that broke the camel's back." The so-called reasons for aggressive driving are actually triggers. In most human behaviour there are different levels of motivation, and aggressive driving is no exception. While the event that sparks the aggression may seem trivial, in every case there exists some stored anger or frustration that is released by the triggering incident (Mizell 1997). It is likely that the cause of the road rage extends beyond the immediate incident. A driver may have had a bad day at work or troubles at home. It may be difficult to tackle the cause of the frustration, and the driver may not even identify feelings of frustration. However, perceived “bad driving” by another driver may be enough to trigger a release of the pent-up frustration (Joint 1995).

It is widely accepted that there are environmental variables that can either provoke aggression or increase the likelihood of its occurrence. Noise is one of them and research suggests that noise influences the intensity of already existing aggression. If an individual has no control over an irritating noise (the volume or duration), it produces stress, makes concentration more difficult, and raises the level of aggression. In congested driving for example, noises produced by other vehicles could increase aggression. Temperature also affects aggression and frustration. Violent crimes increase during summer months, even though data that confirms the link between temperature and aggression is sparse.
In a controlled environment where the influence of temperature on aggressive driving behaviour was studied, the results show a correlation between temperature and aggression (Kenrick and MacFarlane 1986). When positioned behind a vehicle that blocked traffic, drivers subjected to higher temperatures, started honking their horns earlier, they also honked their horns more frequently and for longer than drivers subjected to lower temperatures.

Another potential explanation to the phenomena of aggressive driving behaviours is in the perception of the car itself combined with overcrowding (another subjective variable). The car is symbolic in many ways, often it is the individual's second most valuable belonging. It is often important for the owner's livelihood, provides access to freedom, and is furthermore also a "statement of self" (Connell and Joint 1996). Its size, shape, power, colour, and value may all be used by the owner as an expression of how they see themselves and how they want others to see them. Every time the car is used its value and meaning is to some extent controlled and obstructed by forces beyond the driver's control, and it is placed at an unknown risk by other road users. Driving is an emotive activity, and the car is a prized and symbolic possession which is uniquely able to provoke personal offense and territorial defence if any perceived threat occurs (Connell and Joint 1996). Human beings are territorial, and anyone who invades that territory is potentially an aggressor. The car is an extension of territory, and the territory extends for some distance beyond the vehicle. If a vehicle threatens this territory by cutting in, for example, the driver will probably carry out a defensive manoeuvre (Joint 1995). The defending driver may also go one step further and assert dominance, and drivers admit to having chased after a driver to "teach him a lesson," often tailgating in the process. Sometimes there is gesticulations and aggressive manoeuvres, this might also degenerate into drivers physically assaulting each other or each other’s vehicles (Joint 1995). Driving may also be a field where stress and tension can accumulate, without providing an outlet. Congestion is also undoubtedly an issue here, where drivers must also adhere to limitations placed on their speed and movement, by road layout, regulations and other drivers.

A study by the AA Foundation (Rolls and Ingham 1992) revealed that one of the main factors influencing driver behaviour was mood. Unsafe drivers were affected by mood to a much larger extent than safe drivers. This might be due to the fact that, for many of the unsafe drivers, the act of car driving is regarded as an expressive, rather than practical, activity. Being in a bad mood appears to have an adverse effect on driving behaviour and this effect appears to be most pronounced among unsafe drivers. Unsafe drivers were more likely to get wound up about what they see as inappropriate or "stupid" actions of other road users. The bad moods of the driver were more likely to be exacerbated by other driver actions (Rolls and Ingham 1992).
In experiments where people are encouraged to vent aggression and then record the emotional results, data suggests that human aggression is not simply an innate drive. If aggression was a basic biological drive it should be cathartic so that after people exhibit aggressive behaviours their anger and frustration should be satiated. Data from experiments show that venting "pent-up" anger by swearing and gesticulating does not resolve the problem (Geen 2001). Venting anger appears to do little or nothing to reduce feelings of aggression. Goleman (Goleman 1997) provides perhaps one of the most accessible scientific explanations for rage (not specifically road rage). As Goleman puts it “...anger is the most seductive of the negative emotions; the self-righteous inner monologue that propels it along fills the mind with the most convincing arguments for venting rage”. Anger is energizing and exhilarating and Goleman suggests that this explains why people believe that anger is uncontrollable (or should not be controlled) and that venting anger is "cathartic" even though research fails to support these beliefs. With anger the limbic system will release catecholamines (an organic compound), which prepares the individual to take flight or fight depending on the situation - what Goleman refers to as the "rage rush" (Goleman 1997). Even though this state will only last for a few minutes, the limbic system also prompts arousal, providing a longer-lasting state of readiness. This state of arousal lowers the threshold for provoking anger (Zillman 1993) so that anger builds on anger and the emotional brain heats up (Goleman 1997). Rage becomes unhampered by reason and can easily erupt into violence. People become unforgiving and beyond being reasoned with and their thoughts revolve around revenge and reprisal.

Proposed interventions to reduce road rage range often involve physical activity and the sooner the intervention the more effective it is. (Geen 2001) Drivers should be advised to find a situation with a low risk for further provocation and wait for the anger/readiness state to wear off. Distraction is a key device in achieving this psychological "cooling off" (Goleman 1997) and would include activities such as long walks, active exercise or specific relaxation methods. Non physical distractions such as TV, films, and reading also aided cooling off by interfering with the anger cycle (Tice and Baummeister 1993). Leading to hopes that appropriately designed in-vehicle systems might achieve the same results.

### 2.2.1 Voice Discrimination and Familiar Voices

People are extremely skilled at recognizing and tuning into a specific voice even when this voice is one of many heard for instance in a room full of people. In an fMRI study Stevens (Stevens 2004) found that a particular brain region was involved in recognizing and discriminating voices. The study where listeners were asked to determine whether two voices were the same or whether two words were the same found that voice comparisons were made in the right frontal parietal area, and word
processing was done in the left frontal and bilateral parietal areas. Other studies found that familiar voices are processed differently than unfamiliar voices, and famous voices are recognized using different regions of the brain than when discriminating between unfamiliar voices (Van Lancker and Kreiman 1987; Van Lancker, Cummings et al. 1988; Van Lancker, Kreiman et al. 1989). Another set of studies show that the linguistic properties of speech, what is actually said, is processed in a region of the brain different from the regions that recognize and discriminate between voices (Kreiman and Van Lancker 1988; Glitsky, Polster et al. 1995). Together these studies show that voice discrimination is processed differently from and distinct from what is said, even though conveyed in the same speech stream.

Still, properties of a voice affect what is said and studies show that consistency and familiarity of voice affect how the message is processed and perceived. A study where a set of words were spoken by either a female or male voice show that reaction times to recognize/categorize the words were slower when two voices were used than when all the words were spoken using one voice (Mullennix and Pisoni 1990). They also found that increasing the number of voices further slowed down the time it took to recognize/categorize the recorded words. Similarly, results from a study that examined the effects of familiarity of voice on recall for spoken word lists, showed that lists produced by multiple voices lead to decreased accuracy recall compared to lists of words produced by a single voice. A study where listeners were asked to type the word they heard presented in noise (Goldinger 1996) found the same trend, words spoken by the same voice were recognized more often than words spoken by different voices. Follow-up studies show that the advantage of familiar voices also holds for sentences (Nygaard and Pisoni 1998). Taken together, these studies show that consistency of voices and familiarity of voices helps both recognition and recall of spoken language.

Famous and familiar voices could potentially have advantages for conveying messages. Famous people, and especially media people, are often trained in how to use their voices and can therefore be better at reading and recording scripts needed for an in-vehicle information system. Both radio and TV presenters are selected in part for their voices. Furthermore, matching a famous or familiar voice to the content of a message might increase credibility and recall of the message (Plapler 1974; Misra and Beatty 1990). A study where a famous voice was compared to unknown voices in an advertising campaign confirm these results even though there was no increase in people’s willingness to buy the product when presented with the famous voice (Leung and Kee 1999). Gender of voices also interacts with the content of the message. A study that tested listener’s attitude towards different products when presented by a female or male voice found that the gender of the presenter’s voice does not affect
gender neutral or male gender products, but has a strong effect on female gender products (Whipple and McManamon 2002). Their results also show that a female voice worked better for female gender products when the intended buyer was female, and that a male voice worked better when the intended buyer was male.

As mentioned in Chapter 1, effects of accents still need research, even though some results have been obtained (Dahlbäck, Wang et al. 2007). Most important for an in-vehicle system in a noisy car environment: accented speech was found to be less comprehensible and harder to process when mixed with noise (Lane 1963; Munro and Derwing 1995; Munro 1998).

The message and topic matters so that voice characteristics will affect the listener more or less dependent on how involved and interested the listener is in the message conveyed by the voice. Studies show effects on both attitude and memory. A study where both the content (interesting or non-interesting), and the voice (high intensity and intonation vs. low intensity and intonation) was varied show that voice characteristics matter when the message initially is not interesting (Gelinas-Chebat and Chebat 2001). Engaging qualities of voice characteristics with intensity and varied intonation grabs the listener’s attention for low-engagement messages. Goldinger (Goldinger 1996) investigated how changes in voices interacted with the focus of the listener’s attention. Similar to the results on voice characteristics and attitude, changes in voice characteristics do not matter when the listener is focused on the meaning of the message. Conversely, when the listener is listening in a shallow manner, changes in voice characteristics have a detrimental effect on recall.

### 2.2.2 Design of Study

Frustration and anger has many unpleasant side effects including increased likelihood of becoming even angrier, decreased ability to pay attention, decreased ability to think and problem solve, and also to interact with others. Driving is an activity where attention, judgment and performance is of great to increase traffic safety (Joint 1995; Connell and Joint 1996; Mizell 1997). The purpose of this study is to investigate the impact anger and frustration on driving safety and in particular to study the impact of familiar and unfamilar voices used in in-vehicle information systems. To do this, we setup a driving simulator experiment were we used a conversational in-vehicle system with two voices, familiar and unfamiliar. In the study we measured how emotionally induced angry and frustrated drivers responded to interaction with the in-vehicle system in terms of emotional state, attitude and driving performance.

The study was designed as a 2x3 study; Gender (Male, Female) x In-Vehicle system (Familiar voice, Unfamiliar voice, No System). For this experiment, the desired
emotion – anger and frustration - was created at the beginning of the experiment rather than to rely on participants being in the required state of mind when arriving to the study or participants acting the emotion (Masters, Jonsson et al. 2006).

2.2.3 Participants

60 participants, 30 female and 30 male, in the age group of 18-25 were recruited from Oxford Brookes University in the UK. All participants had drivers’ licenses and were native English speakers. Participants were paid for their time.

2.2.4 Procedure

All participants were informed that the experiment would take 1.5 hours and started the experimental session by signing the consent form. This was followed by the first set of questionnaires with general information such as gender, age and driving experience.

Following this each participant went through a 5 minute introduction to the driving simulator setup. This consisted of a commercial driving simulator – STISIM Drive - a car seat and a Microsoft Sidewinder steering wheel and pedals consisting of accelerator and brake. This introduction also included a 3 minute driving course to familiarize participants with the simulator and screen for simulator sickness (Bertin, Guillot et al. 2004). One participant felt nauseas and did not complete the experiment; another participant was recruited as a replacement.

Participants were then induced to be angry and frustrated using three methods taking in total 45 minutes. The reason for these four methods is that anger is one of the more difficult emotions to induce and several researchers have reported difficulty eliciting high levels of reported anger (Phillipot 1993; Gross and Levenson 1995).

First all participants viewed a 15 minute compilation of the movie Cry Freedom. The reason for selecting a video is that anger is one of the more difficult emotions to elicit with film clips. Longer time is needed to induce anger and often used films; “My Bodyguard” and “Cry Freedom”, all revolve around themes of injustice. Injustice will elicit a response of deep anger in most people.

Second, all participants’ were subject to two computer based tasks. These tasks were based on techniques that induce anger through interpersonal or interactive situations. The reason why these type tasks elicit anger and frustration is that anger requires a high level of personal engagement and/or immediacy. While this is hard to achieve with a film, it is easy with tasks that require timely responses and interactions. Two
interactive sessions were designed where participants interacted with two computer based tasks:

- Participants started with a Stroop colour-naming task. This task was first demonstrated by J. R Stroop in 1935 (Stroop 1935) This task is based on three classes of stimuli, for instance: congruent – red is written in red ink, conflicting: red is written in green ink, and control: red written in black ink. The main effect here is that word information interferes with the naming of colours, while colour information does not interfere with the naming of the word. This is called the Stroop effect. Reaction times are such that colour naming always takes longer than word reading, however, in the conflicting case, reaction times to colour naming is significantly higher. (Stafford 2003). Participants were presented with a sequence of pages with 5 words on each page. Each word was a colour (red, blue etc) and the colour word was presented in a font with a different colour ink. Participants were asked to say out loud the colour of the ink that each word was written in. Pages were presented for a brief period of time, enough time to read 5 words or say 5 colours. Please note that this particular task has to be done in a language that participant’s know well to get the desired conflicting effect.

You will see a number of words written in colour on the following screens. For each screen, please say out loud the colours the words are written in.

(Click to start)

Blue is printed in red, red is printed in yellow, yellow in purple, green in orange, and orange in green. The Epress version is printed in colour.

- After the Stroop task, participants did an arithmetic task where they were asked to count backwards in 7s from a four digit number. With regular interval, participants were presented with a new page containing a new start number in the sequence and asked to continue subtracting 7s.
Finally as the last task, the inducement session was concluded by participants going through a visualization task (Fessler, Pillsworth et al. 2004) During this task participants were asked to recall or imagine a time when they had experienced anger and frustration, and then to write a brief essay about that time. In particular “Imagine that someone has done something to make you really angry. Briefly describe the circumstances that would make you the angriest”. “Jot down, as specifically as you can your feeling and emotions in response to the angry situation you just described”.

A questionnaire, the Differential Emotional Scale – DES - (Izard 1977), where participants self-reported on emotional state was administered before and after the inducement session as the manipulation check.

After the emotional inducement participants were randomly divided into three gender balanced groups of twenty. All participants drove the driving simulator with specially designed driving scenario that included several hazards within a varied and realistic road scenario. The driving course was 52,000 feet (15.85 kilometres) long and divided into four equal length segments. The driving course was segmented so that driving behaviour could be tracked over time. There were a total twelve road hazards and four traffic events, spread equally over the 4 segments. These hazards and traffic events where prompted by 32 hazard and warning prompts so that each segment of the road was instrumented with eight prompts. Prompts for each segment had the following properties: three prompts that acknowledge bad driving conditions, four prompts designed to elicit interaction, either as questions or by self-disclosure, and one suggestion prompt.

A segment of the driving course typically consisted of the road and traffic layout shown in Figures 2-1 and 2-2.
Please note drivers drive on the left side of the road since the study was done in the UK.

Depicted here is a typical situation where the driver was given a prompt on bad driving conditions. In this case the driver is warned that in the particular area, shopping district, pedestrian normally cross the road without looking.

The driving scenario furthermore included elements with complex and frustrating traffic events to sustain the induced emotion, anger and frustration. These traffic events consisted of situations with heavy traffic, accidents blocking lanes, pedestrians crossing the street without looking, and weather.

For the driving simulator session, drivers sat in a car seat and drove using a Sidewinder steering wheel and pedals. The simulated journey was projected on a wall in front of participants. All drivers completed the same driving scenario since a driving scenario in STISIM Drive is static and predetermined; it has a specific length and will take all drivers along the exact same road regardless of left and right turns. Based on this feature of STISIM Drive all participants are guaranteed to drive the exact same route.

Of the three groups of drivers, one group was driving in silence – no speech based in-vehicle system. The other two groups of 20 were driving with an in-vehicle information system. The information provided in some of the prompts was of an...
Driver Emotion and Properties of Voices

ephemeral type that could feasibly be supplied by police reports or weather reports. There were also prompts based on more permanent information such as the location of school zones and speed limits. Finally, some of the prompts were phrased as questions to initiate a dialogue with the driver. There were in total twelve prompts that acknowledged bad driving conditions throughout the driving scenario. Sixteen prompts were designed to elicit interaction, either as questions or by self-disclosure followed by a question, and four prompts contained information about the road or traffic situation followed by a suggestion to the driver.

Listed below are sample prompts from all three types of messages:

- **Acknowledgement**
  - This road is very busy during rush hour, and traffic is slow.
  - The stop sign in this intersection slows down traffic.
  - This is a really windy stretch of road.

- **Questions – Self-Disclosure**
  - I get stressed in traffic almost every day, how often are you stressed by traffic problems?
  - I like to drive with people that talk to me, what is your favourite person to drive with?
  - Do you find it stressful talking to people while driving?

- **Suggestions**
  - There is an accident ahead; if you turn left here you might avoid it.
  - Traffic is really heavy, if you turn left you might avoid some of it

A complete list of prompts can be found in Appendix A. All prompts were recorded in a neutral female voice using an unknown female voice talent and a familiar female voice talent. Familiarity was guaranteed since one of the voice talents was a well liked lecturer at the University. Of the two groups that were driving with the in-vehicle system, one group drove with the in-vehicle system with the familiar voice, and the other group drove with the in-vehicle system with the unknown voice.

After the driving session, participants self reported on their emotional status using a DES (Izard 1977), the same questionnaire that was used as a manipulation check. They also reported on their feelings trust and liking of the system and car, and on their driving experience in a set of questionnaires. Participants were furthermore asked if they knew the familiar voice and they could all name the lecturer.
2.2.5 Measures and Dependent Variables

2.2.5.1 Prior Driving Experience
Information about participants’ real life driving experience was collected as part of the first questionnaire. All participants were asked about prior driving experience including where they normally drive – city, motorway or rural, how many miles driven per week, and how many accidents and tickets they had had.

2.2.5.2 Manipulation Check – Emotional Inducement
The effectiveness of the emotional inducement was checked by having participants’ self-report using a questionnaire with a 32 term DES (Izard 1977). The index that measured how angry and frustrated participants were was based on the question, “How well do each of the following adjectives describe how you feel?” followed by a list of adjectives based on ten-point Likert scales (1=Describes Very Poorly to 10=Describes Very Well). The index was comprised of 12 items, Sad, Upset, Distressed, Unhappy, Dislike, Distaste, Disgust, Repulsion, Hostile, Angry, Mad, and Aggressive. The index was very reliable (Cronbach’s α = .90).

2.2.5.3 Driving performance
The driving simulator automatically keeps a log of participants driving behaviour, such as accidents, speeding and adherence to traffic regulations. Three indices were created, and index “Accidents” a combination of Off-road accidents and Collisions, and index “Breaking Traffic-Rules” as a combination of Speeding tickets, Traffic light tickets, and Stop sign tickets, and an index for “Bad Driving” as a combination of six items: Off-road accidents, Collisions, Speeding tickets, Traffic light tickets, Stop sign tickets, and Swerving.

2.2.5.4 Driving Experience
The driver’s were asked about various aspects of their driving experience in the post-driving questionnaire, and two indices were created. The first index was based on the question “How well do each of the following adjectives describe how you felt while driving with the in-vehicle system?” The questions was followed by a list of adjectives based on a ten-point Likert scales ranging from Describes Very Poorly (=1) to Describes Very Well (=10). The index, Perceived Emotional Influence, was comprised of nine adjectives, calm, at ease, comfortable, self confident, relaxed, secure and tense, nervous and confused reverse coded. The index was very reliable (Cronbach’s α = .91).

---

\(^1\) Cronbach’s α is explained in section 1.4.5
The second driving experience index, *Perceived Attentiveness*, was based on the question, “How well do the following statements describe the in-vehicle systems influence on your driving?” The questions was followed by a list of statements based on a ten-point Likert scales ranging from Describes Very Poorly (=1) to Describes Very Well (=10). The index was comprised of six terms, Alert driver, Careful driver, Safe driver, Confident driver and Distracted and Aggressive driver reverse coded. The index was very reliable (Cronbach’s α=.88).

### 2.2.5.5 Driver’s Assessment of the in-vehicle system and Car

Driver’s assessed the quality of the in-vehicle system and the car; this was measured by indices created from a questionnaire with the question “How well do each of the following adjectives describe how you feel about the in-vehicle system?” The questions was followed by a list of adjectives and terms based on a ten-point Likert scales ranging from Describes Very Poorly (=1) to Describes Very Well (=10). The index for *Quality of the in-vehicle system* was comprised of nine adjectives and terms: Trustworthy, Friendly, Intelligent, High quality, Helpful, Reliable, and Annoying, Frustrating and Condescending reverse coded. The index was very reliable (Cronbach’s α=.92). The index for *Quality of car* was comprised of nine adjectives and terms: Fun, Well-designed, Want-to-have, High-quality, Likable, Useful, Helpful, Trustworthy and Recommendable. The index was very reliable (Cronbach’s α=.90).

### 2.2.6 Results

The effects of the driver’s emotion and the familiarity of the voice used by the in-vehicle system were measured by paired-samples t-tests and a two-way ANOVA, with gender and in-vehicle system as between-participants factors. When appropriate, a subset of the data was analyzed using a two-way ANOVA. Here the two voice conditions were included and the silent condition excluded. Significance levels are not indicated in the tables, but rather in the text. I am reporting results from significant levels of .05. Note that M is short for Mean and SD is short for Standard Deviation.

#### 2.2.6.1 Prior Driving Experience

The data show some fluctuations but no significant differences in prior driving experience between the 3 participant groups. When analyzed with a two-way ANOVA there were no significant main effects or interaction effects between the groups on accidents, $F(2,54) = .85, p < .43$, on tickets, $F(2,54) = .97, p < .38$, driving with passenger, $F(2,54) = .39, p < .68$, for driving in rush hour, $F(2,54) = 1.0, p < .37$, or for being upset while driving $F(2,54) = 1.8, p < .18$, see Table 2-1.
2.2.6.2 Manipulation Check

To measure the effect of inducement, all participants self-reported on emotional status before inducement and after inducement. The “Angry and Frustrated” index showed similar means, and no significant difference in emotional state between the conditions before the inducement, $F(2,54) = .34$, $p < .71$, see Table 2-2.

Table 2-1: Prior Driving Experience

<table>
<thead>
<tr>
<th>Driving Experience</th>
<th>Value range</th>
<th>Condition</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accidents</td>
<td>Absolute value: number of accidents</td>
<td>Silent</td>
<td>.90</td>
<td>1.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Familiar voice</td>
<td>.60</td>
<td>.88</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Unfamiliar voice</td>
<td>.42</td>
<td>.75</td>
</tr>
<tr>
<td>Tickets</td>
<td>Absolute value: number of tickets</td>
<td>Silent</td>
<td>.90</td>
<td>1.3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Familiar voice</td>
<td>.40</td>
<td>1.2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Unfamiliar voice</td>
<td>.06</td>
<td>.22</td>
</tr>
<tr>
<td>Driving with Passengers</td>
<td>Maximum Value of 3, for always driving with passengers, and 0 for never driving with passengers</td>
<td>Silent</td>
<td>2.0</td>
<td>1.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Familiar voice</td>
<td>1.9</td>
<td>.64</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Unfamiliar voice</td>
<td>2.1</td>
<td>.97</td>
</tr>
<tr>
<td>Drives in Rush Hour</td>
<td>0 for no, 1 for yes</td>
<td>Silent</td>
<td>.75</td>
<td>.44</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Familiar voice</td>
<td>.75</td>
<td>.44</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Unfamiliar voice</td>
<td>.60</td>
<td>.50</td>
</tr>
<tr>
<td>Gets upset while driving</td>
<td>0 for no, 1 for yes</td>
<td>Silent</td>
<td>.35</td>
<td>.49</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Familiar voice</td>
<td>.50</td>
<td>.51</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Unfamiliar voice</td>
<td>.50</td>
<td>.51</td>
</tr>
</tbody>
</table>

The data also clearly show that the “Angry and Frustrated” index is affected by the inducement, using a paired samples t-test, $t(59) = 34.5$, $p < .001$. Means of the index was $M=15.96$, $SD=4.2$ before inducement, and $M=57.38$, $SD=7.6$, after inducement.

Table 2-2: Manipulation Check, emotional state before and after inducement

<table>
<thead>
<tr>
<th>Conditions</th>
<th>Gender</th>
<th>Significance level</th>
<th>Before inducement Mean</th>
<th>SD</th>
<th>After Inducement Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Silent</td>
<td>Female</td>
<td></td>
<td>15.25</td>
<td>3.50</td>
<td>55.68</td>
<td>5.06</td>
</tr>
<tr>
<td></td>
<td>Male</td>
<td></td>
<td>14.90</td>
<td>3.60</td>
<td>55.60</td>
<td>6.98</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>P &lt; .001</td>
<td>15.08</td>
<td>3.46</td>
<td>55.64</td>
<td>5.98</td>
</tr>
<tr>
<td>Familiar Voice</td>
<td>Female</td>
<td></td>
<td>15.47</td>
<td>3.18</td>
<td>57.87</td>
<td>6.46</td>
</tr>
<tr>
<td></td>
<td>Male</td>
<td></td>
<td>16.57</td>
<td>3.82</td>
<td>56.93</td>
<td>6.15</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>P &lt; .001</td>
<td>16.02</td>
<td>3.47</td>
<td>57.40</td>
<td>6.16</td>
</tr>
<tr>
<td>Unfamiliar Voice</td>
<td>Female</td>
<td></td>
<td>17.32</td>
<td>6.02</td>
<td>58.19</td>
<td>7.30</td>
</tr>
<tr>
<td></td>
<td>Male</td>
<td></td>
<td>16.24</td>
<td>4.68</td>
<td>60.01</td>
<td>12.44</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>P &lt; .001</td>
<td>16.78</td>
<td>5.28</td>
<td>59.10</td>
<td>9.97</td>
</tr>
</tbody>
</table>
There are no significant differences in emotional state between driving conditions after inducement, $F(2,54) = .17, p < .85$.

### 2.2.6.3 Emotional State after Driving

The data on emotional state collected after the driving session show that there are significant differences between the emotional states of participants that drove in different conditions. There is a main effect for condition. Drivers with a familiar voice feel the best after driving, and drivers with the unfamiliar voice feel the worst, with the silent condition in the middle, $F(2,54) = 3.8, p < .03$, see Table 2-3.

<table>
<thead>
<tr>
<th>Conditions</th>
<th>Gender</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Silent</td>
<td>Female</td>
<td>30.90</td>
<td>8.11</td>
</tr>
<tr>
<td></td>
<td>Male</td>
<td>31.58</td>
<td>11.30</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>31.24</td>
<td>9.58</td>
</tr>
<tr>
<td>Familiar Voice</td>
<td>Female</td>
<td>20.06</td>
<td>6.18</td>
</tr>
<tr>
<td></td>
<td>Male</td>
<td>31.37</td>
<td>11.21</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>25.71</td>
<td>10.55</td>
</tr>
<tr>
<td>Unfamiliar Voice</td>
<td>Female</td>
<td>38.64</td>
<td>16.74</td>
</tr>
<tr>
<td></td>
<td>Male</td>
<td>31.96</td>
<td>10.07</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>35.30</td>
<td>13.87</td>
</tr>
</tbody>
</table>

There is no main effect for gender, $F(1, 54) = .38, p < .54$. There is however an interaction effect between gender and condition, $F(2,54) = 3.3, p < .05$, see Table 2-3. Female drivers are more affected by familiarity of voice than male drivers. Female drivers with a familiar voice felt significantly less angry and upset after driving than female drivers with the unfamiliar voice. The data show no such effect for male drivers, the means in conditions show very little difference.

### 2.2.6.4 Driving performance

There data showed no differences in driving behaviour (accidents, speed and adherence to traffic regulations) between the four segments of the driving course. The driving session was hence treated as one continuous drive, instead of four segments. Analyzing the driving course as one segment, the data show clear differences between conditions for all three indices of driving performance, accidents, adherence to traffic regulations, and the combined index of “bad driving” that includes accidents, breaking traffic regulations and swerving.

The data show that participants driving with the familiar voice had fewer accidents than drivers with the unfamiliar voice $F(2, 54) = 3.5, p < .04$. There were no gender
Social and Emotional Characteristics of Speech-based In-Vehicle Information Systems: Impact on Attitude and Driving Behaviour

The in-vehicle voice system also affected adherence to traffic regulations. Drivers with the familiar voice were the most law-abiding, and drivers without the system broke the most traffic rules, $F(2,54) = 3.2, p < .05$. There were no gender differences, $F(1,54) = 2.1, p < .16$, or interaction effects, $F(2,54) = .30, p < .75$, see Table 2-4.

Not surprisingly, the data also show the same trends for the combined index of driving behaviour as for the more specific behaviours. Drivers with the familiar voice did significantly better than both the unfamiliar voice and without a voice system, $F(2,54) = 3.24, p < .05$. Female drivers drove better across conditions, $F(1,54) = 4.70, p < .04$, than male drivers. There were no interaction effects, $F(2,54) = .55, p < .60$, see Table 2-4.

A comparison between the two groups driving with a voice system in the car shows a significant improvement in driving with the familiar voice over the unfamiliar voice. The effect for accidents is $F(1,36) = 6.8, p < .02$, the effect for traffic regulations is $F(1,36) = 3.8, p < .05$, and the effect for the combined index of bad driving is $F(1,36) = 5.0, p < .03$, see Table 2-4.

There are no significant differences for any of the driving behaviour measures between the group driving with the unfamiliar voice and the group driving without the voice system. The data for accidents show no difference, $F(1,36) = .81, p < .38$, neither does the data for traffic rules $F(1,36) = .21, p < .65$, nor the measure for the overall combined driving behaviour $F(1,36) = .06, p < .81$, see Table 2-4.

Table 2-4: Driving Performance – Bad Driving

| Conditions       | Gender | Accidents | | Breaking Traffic Rules | | Overall Bad Driving |
|------------------|--------|-----------|-----------------|---------------------|-------------------|
|                  |        | Mean | SD | Mean | SD | Mean | SD |
| Silent           | Female | 5.4  | 2.01 | 3.5 | 2.50 | 39.40 | 11.85 |
|                  | Male   | 6.0  | 2.62 | 5.0 | 2.21 | 41.89 | 9.75  |
|                  | Total  | 5.7  | 2.39 | 4.3 | 2.41 | 40.65 | 10.64 |
| Familiar Voice   | Female | 3.4  | 1.26 | 2.2 | 1.23 | 27.84 | 8.60  |
|                  | Male   | 5.6  | 1.96 | 3.0 | 2.40 | 38.10 | 10.48 |
|                  | Total  | 4.5  | 1.96 | 2.6 | 2.10 | 32.97 | 10.71 |
| Unfamiliar Voice | Female | 6.4  | 2.95 | 3.8 | 2.10 | 38.12 | 8.62  |
|                  | Male   | 6.5  | 2.88 | 4.1 | 2.73 | 45.14 | 18.37 |
|                  | Total  | 6.4  | 2.84 | 4.0 | 2.37 | 41.63 | 14.43 |
2.2.6.5 Driving Experience and Attitude

There were two indices concerned with the drivers driving experience and attitude towards the in-vehicle system. Drivers perceived the in-vehicle system with the familiar voice to be a positive influence, they claim it made them more safe and confident as drivers than the system with the unfamiliar voice, \( F(1,36)=10.8, p < .002 \), there was no main effect for gender, \( F(1,36) = .04, p < .85 \) and no interaction, \( F(1,36) = .04, p < .85 \), see Table 2-5.

Drivers that drove and interacted with the in-vehicle system with the familiar voice rated themselves as significantly more attentive while driving, \( F(1,36) = 15.0, p < .001 \), than drivers interacting with the unfamiliar voice, see Table 2-5. There was no main effect for gender, \( F(1,36) = .68, p < .42 \), or interaction, \( F(1,36) = 2.2, p < .15 \).

<table>
<thead>
<tr>
<th>Indices</th>
<th>Gender</th>
<th>Familiar Voice</th>
<th>Unfamiliar Voice</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
</tr>
<tr>
<td>Perceived Positive Influence of system</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>55.70</td>
<td>15.0</td>
<td>41.04</td>
</tr>
<tr>
<td>Male</td>
<td>55.71</td>
<td>16.5</td>
<td>39.19</td>
</tr>
<tr>
<td>Total</td>
<td>55.71</td>
<td>15.3</td>
<td>40.12</td>
</tr>
<tr>
<td>Perceived Attentiveness as drivers</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>35.80</td>
<td>9.6</td>
<td>21.30</td>
</tr>
<tr>
<td>Male</td>
<td>29.53</td>
<td>7.5</td>
<td>23.10</td>
</tr>
<tr>
<td>Total</td>
<td>32.67</td>
<td>9.0</td>
<td>22.20</td>
</tr>
</tbody>
</table>

2.2.6.6 Assessment of In-Vehicle System and Car

The perceived quality of the in-vehicle system was clearly affected by the voice of the system. Drivers driving with the familiar voice in the in-vehicle system rated the quality of the in-vehicle system much higher than drivers with the unfamiliar voice, \( F(1,36) = 31.2, p < .001 \), see Table 2-6.

<table>
<thead>
<tr>
<th>Indices</th>
<th>Gender</th>
<th>Familiar Voice</th>
<th>Unfamiliar Voice</th>
<th>No Voice</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
<td>SD</td>
</tr>
<tr>
<td>Quality of in-vehicle system</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>60.22</td>
<td>9.0</td>
<td>36.6</td>
<td>7.75</td>
</tr>
<tr>
<td>Male</td>
<td>45.87</td>
<td>10.2</td>
<td>36.44</td>
<td>10.3</td>
</tr>
<tr>
<td>Total</td>
<td>53.05</td>
<td>11.9</td>
<td>36.5</td>
<td>8.9</td>
</tr>
<tr>
<td>Quality of car</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>50.96</td>
<td>8.0</td>
<td>33.48</td>
<td>12.0</td>
</tr>
<tr>
<td>Male</td>
<td>30.57</td>
<td>8.9</td>
<td>30.62</td>
<td>12.1</td>
</tr>
<tr>
<td>Total</td>
<td>40.76</td>
<td>13.3</td>
<td>32.05</td>
<td>11.8</td>
</tr>
</tbody>
</table>

There was a gender difference where female drivers rated the quality of the in-vehicle system higher than male drivers, \( F(1,36) = 6.0, p < .02 \). The data also showed an
interaction effect. Female drivers with the familiar voice rated the quality of the in-vehicle system much higher, $F(1,36) = 5.7, p < .02$, than male and female drivers with the unfamiliar voice, with male drivers with the familiar voice in the middle, see Table 2-6.

Drivers perceived the quality of the car according to the same pattern as the quality of the in-vehicle system. Drivers with the familiar voice in the in-vehicle system rated the quality of the car higher than drivers with the unfamiliar voice-system in the car, or drivers without a system, $F(2, 54) = 5.2, p < .009$. There was also a main effect for gender. Female drivers rated the quality of the car higher than male drivers, $F(1,54) = 11.5, p < .001$. There was also an interaction effect, where female drivers with the familiar voice rated the quality of the car much higher than male drivers with the familiar voice and also higher than all other conditions, $F(2,54) = 3.3, p < .04$, see Table 2-6.

2.2.7 Discussion: Familiarity of Voice for Angry and Upset Drivers

It was interesting to note that the inducement for anger and frustration seemed to hold for the duration of the driving session. The reason for dividing the driving session into four segments was to be able to track driving behaviour over time, from the first 5 minutes after inducement to about 20 minutes after inducement. Anger is not an emotional state that is desirable or comfortable, and when induced (over only 45 minutes) I feared that very little stimuli would be needed to change the emotional state, unless it was somehow sustained. The driving course was hence setup with frequent frustrating hazards and traffic events, evenly spread over the course. There was no noticeable difference in driving behaviour between the four road segments, the number of accidents and disobeyed traffic regulations were evenly spread over the course, so the entire driving course could be analyzed as one driving session.

Using a familiar voice in an in-vehicle system for angry and frustrated drivers had an overall positive effect. The data show no adverse effects of a familiar voice, instead consistently on all measures, the familiar voice is perceived more positively and has a more positive influence on driving behaviour than the unfamiliar voice.

The familiar voice made drivers feel better after driving compared with the drivers that drove with an in-vehicle system with an unfamiliar voice, the “feeling bad” index show $M = 25.7$ for familiar voice and $M = 35.3$ for unfamiliar voice (see Table 2-6). Furthermore, a familiar voice in the car also had a positive influence on all three driving behaviour measures compared with an unfamiliar voice. There were 30% fewer accidents driving with the familiar voice than driving with the unfamiliar voice, see Table 2-4. Drivers with the familiar voice followed traffic regulations 35% better
than drivers with the unfamiliar voice. For overall driving behaviour including accidents, traffic rules, and lane keeping - drivers with the familiar voice exhibit a 20% better performance than drivers with an unfamiliar voice (see Table 2-4).

Driving experience was also perceived better with a familiar voice in the in-vehicle system. The data show that drivers rated the positive influence of the system at $M = 55.7$ and the attentiveness to the system and the driving task at $M = 32.7$ for the familiar voice, with $M = 40.1$ and $M = 22.2$ respectively for the unfamiliar voice (see Table 2-5). The quality assessment show similar figures with the in-vehicle system with the familiar voice considered being considered 45% better than the system with the unfamiliar voice. Similar, the quality of the car with the in-vehicle system with the familiar voice was considered 30% better than the car with the unfamiliar voice in the in-vehicle system (see Table 2-6). Even though the data show significant improvements of a familiar voice for all drivers, the effect is much stronger for female drivers than male drivers. The reason for this could be the gender bias (the use of a female voice) as discussed below in Chapter 3.

The interesting result here is that the simple approach of using a familiar voice in an in-vehicle system has a consistent, clear and positive effect on driving behaviour and attitude. The familiar voice had a clear positive influence on drivers’ emotional state; they were less angry and frustrated than drivers with the unfamiliar voice at the end of the driving session. Paying attention to a familiar voice, what it says and also to interact with it seemed more interesting and less distracting than interacting with an unfamiliar voice. The result show that listening to and conversing with the unfamiliar voice was more distracting, leading to poorer driving performance with more accidents and less attention to traffic regulations. The spillover/halo effect, assessment and rating of in-vehicle system influence the perception of the car, so that the perception of quality of the in-vehicle system was also transferred to the car.

The results from the study suggest that angry and upset people find it easier to pay attention to and interact with a familiar voice; in this case a familiar voice with positive connotations. This furthermore means that one voice does not fit all, since a voice familiar to one individual can be unknown to another.

There are some possible explanations for why the familiar voice has a positive impact on both attitude and behaviour. The literature show similar results for non emotional people, a known voice has more credibility (Plapler 1974; Misra and Beatty 1990) and also a better potential to be heard and listened to than an unfamiliar voice (Leung and Kee 1999). Drivers would then willingly listen to and interact with the familiar voice, feeling trust and liking. This would explain why they felt better – less angry and
Social and Emotional Characteristics of Speech-based In-Vehicle Information Systems: Impact on Attitude and Driving Behaviour

frustrated - and hence also why they performed better (people in a good mood perform better than people in a bad mood). Another explanation is that the use of a familiar voice in an in-vehicle system might influence the drivers’ perception of anonymity. This loss of anonymity would then have a positive impact on driving behaviour (Ellison, Govern et al. 1995; Ellison-Potter, Bell et al. 2001). Studies show that drivers that are shielded from the outside environment by a roof and tinted windows develop a sense of anonymity and detachment (Ellison, Govern et al. 1995). Some people feel less constrained in their behaviour when they are not monitored, and can exhibit antisocial behaviours (Stuster 2004). The result of being less anonymous more pronounced aggressive driving behaviours. Results from two studies show quicker onset of aggression and longer durations of aggressive behaviours such as honking and tailgating (Stuster 2004). If a familiar voice in an in-vehicle system reduces the drivers feeling of anonymity, this might also show up as a reduction in aggressive and unwanted driving behaviours, leading to fewer accidents and better adherence to traffic regulations. It is possible that anonymity or rather loss of anonymity, can help explain results from this study.

Still remaining though, are questions: how to select the appropriate familiar voice for each individual driver? The voice has to have the right positive connotations and properties for each driver. Furthermore, what makes a voice familiar? How familiar should a voice be – family – authoritative figure – friend? When does a voice shift from being unfamiliar to familiar? Can a voice used in the car lose its good connotations due to mishaps by the system, so that a previously trusted familiar voice becomes a disliked familiar voice? What will be the effect of this in car? And more importantly, what would be the effect in real life? Can this result in liked person becoming disliked? The opposite is of course also possible where a person whose voice is used in the car, makes some blunders that shifts the associations from positive to negative. Will this affect drivers’ attitude and performance?

2.3 Matching Driver Emotion with Emotion of Voice

2.3.1 Emotion and Voices

People respond to characteristics of human voices because they manifest emotions, personality, gender, and accents. This is also true for synthetic voices. These responses are both momentary (short-lived states) and unvarying (traits) aspect of a person. Traits predict a wide range of behaviours and attitudes that allows people to predict a speaker’s behaviour. It is easy to define interaction based on traits; by listening to how people speak and express themselves, you can classify them. Subsequent interactions are then affected by the interpretation of paralinguistic cues such as the rising tone of a question, the staccato of anger, or the familiarity of the
voice a someone that you know. In addition to traits, people are however also affected by their environment. To predict a person’s behaviour at any given moment in time also requires attention to the user’s state, that is, the particular feelings, knowledge, and physical situation of the person at a particular point in time. Emotions are the most powerful of the momentary states (Brave and Nass 2002), and affective states such as short-lived emotional states and longer-term moods influence people when they talk to people, interact with computers, or drive a car.

Interaction with computer based device has traditionally been viewed as an exception where people must shed their emotions to interact efficiently and rationally with these non-human devices. Studies on psychology and technology, however, suggest a different view of the influence of emotions in the interaction between people and technology (Gross, 1999). Some psychologists also argue that it is impossible for a person to perform an action without engaging his or her emotional system (Picard, 1997). So, even if traits mostly guide decisions and choices that people make, it is also true that momentary states, such as emotions, influence these choices and decisions. This is true for both parties in a dialogue, both the speaker and listener.

With potentially more complex devices for the driver to control, using speech no longer becomes a gimmick but more a necessity. Introducing speech-based interaction and conversation into the car, however, highlights the potential influence of linguistic and paralinguistic cues. These cues play a critical role in human—human interactions, manifesting among other things, personality and emotion (Nass and Gong 2000; Tusing and Dillard 2000). Cues such as loudness, fundamental frequency, frequency range, and speech-rate distinguish dominant from submissive individuals and affect people interacting with systems using computer-generated speech (Manstetten, Krautter et al. 2001; Strayer and Johnston 2001). It has also been shown that most basic emotions or modes are associated with acoustic properties in a voice. For example, sadness is generally conveyed by slow and low-pitched speech, while happiness is associated with fast and louder speech (Murray and Arnott 1993; Brave and Nass 2002). The emotional characteristics of the voices used by in-vehicle information systems might thus impact attention, performance, and judgment in same way as the emotion of the driver. It thus becomes important to know if the emotion of the driver and the emotion of the voice of the in-vehicle information system interact to influence the driver’s attitude and performance.

2.3.2 Design of Study

To investigate questions on 1) how the emotional state of a driver impacts driving performance, and 2) how the voice used in an in-vehicle information system interacts
In this driving-simulator-based study we investigated what happened when an upset driver encountered a happy voice in the car, and what happened when an upset driver encountered an upset voice in the car (Nass, Jonsson et al. 2005). To assess the effects of the link between driver emotion and in-vehicle system voice emotion on drivers’ performance, the number of accidents that drivers were involved in during their time in the simulator was recorded. In addition to this, the driver’s attention to the driving task was measured as a reaction time to honk the horn in response to randomly occurring honks. Furthermore, the driver’s engagement with the system was measured by the amount of time drivers spent talking back to the voice in the car while driving down the simulated road.

The experiment was a 2x3 between participant design with Driver Emotional State (Happy or Sad) x In-Vehicle System (Happy Voice, Sad Voice, No system). To address the questions of the effects of driver emotion on driving performance, it was necessary to have half of the participants be happy at the time of the experiment and the other half be sad. Faced with the same three choices of screening and scheduling emotional participants as the previous study, work with participants current emotions, have participants act emotions, or induce emotions, I once again elected to create emotions (Masters, Jonsson et al. 2006). Therefore, either a happy or sad state was induced in each participant at the beginning of the experiment.

2.3.3 Participants

60 participants were recruited in the age group of 18-25. Gender was balanced across conditions so that 30 female and 30 male, participated in the study. All participants were recruited from a temporary agency; they had all had drivers’ licenses for over five years and were native English speakers. Participants were paid for their time.

2.3.4 Procedure

All participants started the experimental session by signing the consent form followed by the first set of questionnaires with general information such as gender, age and driving experience.

Participants were then induced to be either happy or sad. This was accomplished in a five minute inducement session. It is easier (and quicker) to create emotions along the valence dimension than to create lasting emotional states along the arousal dimension (Masters, Jonsson et al. 2006). To create happy and sad emotions, participants were shown 37 six-second film and television clips derived from the
Detenber collection. This inducement session took 5 minutes in comparison to the 45 minute inducement sessions necessary to create anger and frustration. The video clips in the Detenber collection are all rated along the dimensions of valence and arousal. They are extensively used for emotion inducement by communication researchers (Detenber and Reeves 1996). For happy participants, all of the videos scored neutral on arousal and positive on valence, reflected by happy themes. For sad participants, all of the videos scored neutral on arousal and negative on valence, reflected by sad or disturbing themes. A questionnaire, DES (Izard 1977), where participants self-reported on emotional state were administered before and after the inducement session as the manipulation check.

After the emotional inducement all participants drove a driving-simulator. The driving simulator consisted of a PlayStation2 running the game “Hot Pursuit”. The game was configured to run a preset course with predefined properties for both driving conditions and car. The preset course took the driver along highways, country roads, and cities with low to medium traffic implemented as a moderate numbers of cars in both directions. The simulator was projected on a six foot diagonal, projection screen and participants sat in a real car seat and controlled the driving simulator using an accelerator pedal, a brake pedal, and a force-feedback steering wheel. All participants drove on the same simulated course for approximately fifteen minutes.

While driving the course, 2/3 of the participants interacted with a conversational in-vehicle information system that tried to mirror the conversation of a passenger. 1/3 of the participants were driving without the conversational system. The voice of the conversational in-vehicle system was a recorded female voice that made light conversation with the driver. The system introduced itself by saying:

\begin{quote}
Hi. My name is Chris and I will be your Virtual Passenger for today. We are going to be driving today on a coastal road, one that I’ve travelled many times before, but one that may be new to you. The trip shouldn’t take too long: ten or fifteen minutes. Let’s get going.
\end{quote}

At thirty-six separate points along the course, Chris the Virtual Passenger either asked questions, “How do you think that the car is performing?”, and “Do you generally like to drive at, below, or above the speed limit?” or made self-disclosures followed by questions, “I like to drive with people that talk to me; can you tell me about your favourite person to drive with?” and “I saw the new Matrix a couple of days ago and it was Ok, what was the last good movie that you have seen?” All voice prompts for Chris the Virtual Passenger were recorded by the same female voice talent in both a happy-energetic voice and a sad-subdued voice. The emotional colouring, happy-
energetic and sad-subdued, of the virtual passenger prompts was screened and verified by 5 judges. Prompts were re-recorded until consensus had been reached by all 5 judges.

1/3 of the happy and 1/3 of the upset participants were randomly selected to drive and interact with the system with the happy-energetic voice. Similarly, 1/3 of the happy and 1/3 of the sad participants drove and interacted with the system with the sad-subdued voice. The remaining 1/3 of the happy and 1/3 of the sad participants drove without the in-vehicle system.

Everyone wants drivers to pay attention to the road while they receive and process information from in-vehicle information systems so these systems need to be both efficient and have a positive influence on the driving experience. To assess how the voice of the in-vehicle information system influenced the driving experience, participants were invited to speak to Chris the Virtual Passenger as much or as little as they wished. After completing the course, participants filled out a set of post-test questionnaires, self-reporting on emotion using a DES (Izard 1977), assessing the emotion of the voice using a DES (Izard 1977) and rating their driving experience.

2.3.5 Measures

2.3.5.1 Prior Driving Experience

Information about participants’ real life driving experience was collected as part of the first questionnaire. Participants were asked to self report how many years they had been driving, the number of miles they typically drive per week, the number of accidents they had had, and the number of tickets they had received.

2.3.5.2 Manipulation Check – Emotional Inducement

The effectiveness of the emotional inducement was checked by having participants’ self-report using a questionnaires with a 32 term variant of the Differential Emotion Scale (Izard 1977). The index that measured how sad and upset participants were and was based on the question, “How well do each of the following adjectives describe how you feel?” followed by a list of adjectives based on five-point Likert scales (1=Describes Very Poorly to 5=Describes Very Well). The index was comprised of fifteen adjectives: Angry, Disgusted, Distaste, Fear, Mad, Revulsion, Sad, Scared, Shocked, Dislike, Unhappy, Astonished, Upset, Frightened, and Distressed. The index was very reliable (Cronbach’s α=.96).

Drivers’ assessment of the voice emotion of the in-vehicle system (Virtual passenger) was measured by an index created from a questionnaire with the question “How well
do each of the following adjectives describe the Virtual Passenger?” The questions was followed by a list of adjectives based on a five-point Likert scales ranging from Describes Very Poorly (=1) to Describes Very Well (=5). The index for a Happy voice was comprised of eight adjectives: Happy, Enthusiastic, Attentive, Delighted, Interested, Curious, Amazed and Amused. The index was very reliable (Cronbach’s $\alpha=.92$).

2.3.5.3 Driving Performance

To assess the effects of the link between driver emotion and the emotion of the systems voice on drivers’ performance, there were two driving performance measures:

Safety is of utmost importance while driving and the measure Number of accidents was manually coded by watching the recordings from the driving simulator. This was scored by two independent judges, there were no disagreements.

Driver’s attention was assessed by determining the driver’s reaction time to a task-relevant stimulus. Specifically, drivers were instructed to honk their horn as soon as they heard a horn honk. There were 19 randomly-placed honks during the driving session. Because horn-honking is relevant to the driving situation, greater speed is associated with greater attention. This is in contrast to secondary-task reaction time experiments, in which greater response time is associated with more attention to the primary task (Reeves and Nass 1996).

2.3.5.4 Driver’s Engagement with Virtual Passenger

Drivers’ willingness to interact with the system was measured by how much they communicated with the Virtual Passenger. The Driver’s engagement with the system was measured by the amount of time drivers spent talking back to the Virtual Passenger while driving down the simulated road. This was done by listening to the audio recordings from the driving sessions and measuring the length of each utterance using Cool Edit Pro (now called Adobe Audition).

2.3.5.5 Driver’s Assessment of the Virtual Passenger and Car

Driver’s also assessed the quality of the system (Virtual Passenger) and the car and this was measured by indices created from a questionnaire with the question “How well do each of the following adjectives describe how you feel about the Virtual Passenger?” The questions was followed by a list of adjectives and terms based on a ten-point Likert scales ranging from Describes Very Poorly (=1) to Describes Very Well (=10). The index for Quality of the System was comprised of nine adjectives and terms: Trustworthy, Intelligent, Fun, Well-designed, Helpful, Likable, Useful,
Reliable, and Recommendable. The index was very reliable (Cronbach’s α=.92). The index for Quality of car was comprised of nine adjectives and terms: Fun, Well-designed, Want-to-buy, Want-to-have, High-quality, Likable, Useful, Reliable, and Recommendable. The index was very reliable (Cronbach’s α=.95).

2.3.5.6 Driving Experience
To assess the driver’s feelings about various aspects of their driving experience, two indices were created based on the post-driving questionnaire. The first index was based on the question “How well do each of the following adjectives describe how you felt while driving with the Virtual Passenger?” The question was followed by a list of adjectives based on a ten-point Likert scales ranging from Describes Very Poorly (=1) to Describes Very Well (=10). The index, Perceived Emotional Influence, was comprised of ten adjectives, safe, calm, at ease, comfortable, self-confident, relaxed, secure and strained, frustrated and nervous reverse coded. The index was very reliable (Cronbach’s α=.93).

The second index, Perceived Attentiveness, was based on the question, “How well do the following statements describe the Virtual Passenger’s influence on your driving?” The question was followed by a list of statements based on a ten-point Likert scales ranging from Describes Very Poorly (=1) to Describes Very Well (=10). The index was comprised of four terms, Alert driver, Careful driver, Safe driver and Confident driver. The index was very reliable (Cronbach’s α=.83).

2.3.6 Results
The effects of the driver’s emotion and the emotional colouring of the Virtual Passenger’s voice were measured by independent-samples t-tests, paired-samples t-tests, and two-way ANOVAs with Driver emotion and In-vehicle System as between participants’ factors. When appropriate, a subset of the data including the voice conditions but excluding the silent condition was analyzed using two-way ANOVAs. The data was also grouped to use an ANOVA to compare voice conditions with the silent condition. Partial $\eta^2$ is used to compare effect sizes.

Reported in the results are effects to a significance level of .05. Trends – that is effects that are of interest but fails to reach significance - are also reported. Note that M is short for Mean and SD is short for Standard Deviation.
2.3.6.1 Prior Driving Experience
The data showed no significant difference between participants in terms of real life driving experience. Number of accidents \( F(2,54) = 1.6, p < .21 \), tickets \( F(2,54) = 1.4, p < .26 \) and miles driver per week \( F(2,54) = 1.0, p < .37 \) was consistent across conditions.

2.3.6.2 Manipulation Check
As expected, participants who saw the upsetting videos were much more upset, \( M=42.6, SD=10.6 \), than were participants who saw the pleasant video, \( M=17.5, SD=3.5 \), based on a two-tailed \( t \)-test, \( t(58)=12.3, p<.001 \).

Consistent with the intention of the voice manipulation by the Virtual Passenger, the Happy voice was perceived to be much Happier, \( M=29.8, SD=5.0 \), than was the sad voice, \( M=15.0, SD=4.0 \), based on a two-tailed \( t \)-test, \( t(58)=10.3, p<.001 \).

2.3.6.3 Behavioural Effects – Driving Performance
Matching the voice of the car to the drivers’ emotions had enormous consequences. Drivers who interacted with voices that matched their own emotional state (Happy and energetic voice for happy drivers and sad and subdued voice for sad drivers) had less than half as many accidents on average as drivers who interacted with mismatched voices! Matched cases had \( M=3.39 \) versus mismatched cased with a \( M=8.95, F(1,36)=9.01, p<.005 \), see Table 2-7.

The effect of matching the voice was so strong that it reduced the difference in accident rate between happy and sad drivers to non-significant, \( F(1,54)=3.0, p>.09 \). There was no effect for emotion of the voice in the in-vehicle system alone, \( F(1,36)=.12, p < .73 \). It is interesting to note that drivers driving without an in-vehicle system end up in the middle for number of accidents. They have more accidents than the matched conditions but fewer than the mismatched conditions. The large number of accidents that drivers in mismatched conditions experienced made it safer to drive without a system than with a system (both matched and mismatched conditions), \( F(1,56)=4.8, p < .04 \), see Table 2-7.

Drivers paired with matched voices also communicated much more with the voice, even though the voice said exactly the same thing in all conditions, \( M=4.05, SD=1.25 \) for mismatched and \( M=5.32, SD=2.00 \) for matched, \( F(1,36)=6.5, p < .02 \). There were no main effects for driver emotion \( F(1, 36) = .004, p < .95 \), or for voice emotion, \( F (1, 36) = .03, p < .87 \). Although drivers who heard emotion-matched voices spoke to the system more, they were nonetheless better able to avoid accidents.
Table 2-7: Driving Performance Results

<table>
<thead>
<tr>
<th>Voice Type</th>
<th>Number of Accidents</th>
<th>Amount of Speaking</th>
<th>Drivers Attention – number of honks</th>
<th>Drivers Attention – response time</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
<td>SD</td>
</tr>
<tr>
<td>Happy and Energetic Voice</td>
<td>1.9</td>
<td>1.6</td>
<td>5.3</td>
<td>2.2</td>
</tr>
<tr>
<td>Sad and Subdued Voice</td>
<td>8.1</td>
<td>7.7</td>
<td>4.1</td>
<td>1.5</td>
</tr>
<tr>
<td>No Voice</td>
<td>2.2</td>
<td>1.4</td>
<td>15.5</td>
<td>6.4</td>
</tr>
<tr>
<td>Total</td>
<td>4.15</td>
<td>6.2</td>
<td>4.5</td>
<td>2.1</td>
</tr>
<tr>
<td>Happy and Energetic Voice</td>
<td>9.8</td>
<td>7.5</td>
<td>4.0</td>
<td>1.0</td>
</tr>
<tr>
<td>Sad and Subdued Voice</td>
<td>4.9</td>
<td>4.2</td>
<td>5.3</td>
<td>1.5</td>
</tr>
<tr>
<td>No Voice</td>
<td>3.9</td>
<td>2.1</td>
<td>12.9</td>
<td>9.0</td>
</tr>
<tr>
<td>Total</td>
<td>5.9</td>
<td>7.6</td>
<td>4.5</td>
<td>1.4</td>
</tr>
</tbody>
</table>

Drivers’ attention to the driving task was measured by how they responded to horn-honks from the system, number of honks responded to, and average response time.

Analyzing the response times to horn honks, the data show no main effects of driver emotion, \( F(1,54) = .14, p < .71 \). There is however clear evidence that the voice matters. The voice emotion alone did not influence how drivers with a system responded to horn-honks, \( F(1,36) = .07, p < .7 \). There were however significant differences between matched and mismatched conditions in terms of response times to horn-honks. First we found that drivers paired with matched voices responded quicker to horn-honks that drivers paired with a mismatched voice, \( M= 5.9, SD=4.11 \) for mismatched and \( M=3.3, SD=1.57 \) for matched, \( F(1,36) = 6.9, p < .01 \). Second, we found that drivers without a system responded much quicker than drivers with a system \( F(1,56) = 15.1, p < .001 \), see Table 2-7.

There were no significant differences in number of honks that drivers responded to, \( F(2,54) = 2.7, p < .08 \). However, drivers without a system, responded to more horn-honks than drivers with the system, \( F(1,56) = 4.3, p < .05 \), see Table 2-7.
2.3.6.4 Attitudinal Effects – Drivers Assessment of Virtual Passenger and Driving Experience

Drivers that heard voices whose emotions were matched and suited to their own emotion rated themselves as significantly more attentive while driving, $F(1,36)=79.1$, $p<.001$. There were no main effects for driver emotion $F(1,36) = .46$, $p < .50$, or for voice emotion $F(1,36) = .004$, $p < .95$, see Table 2-8.

Participants also perceived that the virtual passenger made them more positive, safer and more confident as drivers in matching conditions than in mismatched conditions $F(1,36)=42.5$, $p < .001$. There were no main effects for driver emotion $F(1,36) = 1.8$, $p < .19$ or voice emotion, $F(1,36) = 1.0$, $p < .32$, see Table 2-8.

Table 2-8: Attitudinal Effects

<table>
<thead>
<tr>
<th></th>
<th>Perceived Emotional Influence</th>
<th>Perceived Attentiveness</th>
<th>Quality of System</th>
<th>Quality of car</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
<td>SD</td>
</tr>
<tr>
<td>Happy Drivers</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Happy and Energetic Voice</td>
<td>73.5</td>
<td>14.8</td>
<td>21.6</td>
<td>5.9</td>
</tr>
<tr>
<td>Sad and Subdued Voice</td>
<td>42.7</td>
<td>7.0</td>
<td>9.2</td>
<td>4.1</td>
</tr>
<tr>
<td>No Voice</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>59.7</td>
<td>18.0</td>
<td>15.4</td>
<td>8.1</td>
</tr>
<tr>
<td>Sad Drivers</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Happy and Energetic Voice</td>
<td>41.3</td>
<td>15.0</td>
<td>8.2</td>
<td>2.8</td>
</tr>
<tr>
<td>Sad and Subdued Voice</td>
<td>63.8</td>
<td>13.3</td>
<td>20.7</td>
<td>4.2</td>
</tr>
<tr>
<td>No Voice</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>50.9</td>
<td>17.2</td>
<td>14.5</td>
<td>7.3</td>
</tr>
</tbody>
</table>

The quality of the voice system was rated higher by drivers that were Happy $M=35.5$, $SD=15$, than those that were Upset, $M=25.5$, $SD=7$, $F(1,36)=22.3$, $p < .001$. Quality was also rated higher by drivers that interacted with the Happy Voice, $M=35.8$, $SD=15.5$, than those that interacted with the Sad Voice, $M=25.2$, $SD=5.3$, $F(1,36)=25.38$, $p < .001$. These effects (Partial $\eta^2 = .38$ and .41 respectively) were overshadowed by the much stronger effect of matching conditions (emotion of the

\[ \text{Partial } \eta^2 \text{ is explained in section 1.4.5} \]
driver matched the emotion of the voice), \( F(1,36)=55.9, p <.001 \), (Partial \( \eta^2 = .61 \)) see Table 2-8.

Drivers’ rating of the quality of the car follows the same pattern as the quality assessment of the system. Quality of the car was rated the highest by matched cases, \( F(1,36)= 27.5, p < .001 \). There was also a main effect of the inducement, where Happy drivers, \( M=46.0 \ SD=15.6 \), rated the quality higher than Sad drivers, \( M=35.7 \ SD=13.9 \), \( F(1,54)=17.7, p<.001 \). The emotional colouring of the voice furthermore influenced the perceived quality of the car (actually more so than the induced driver emotion, Partial \( \eta^2 = .41 \) vs. Partial \( \eta^2 = .25 \)). It is interesting to note that a car with a sad voice was rated higher, \( M=50.6 \ SD=9.6 \), than a car with a Happy voice, \( M=39.0 \ SD=21.0 \), and the car without a voice has the lowest rating \( M=32.83 \ SD=6.2 \), \( F(2, 54)= 18.4, p<.001 \).

### 2.3.7 Discussion – Matching Emotion of Driver and Voice

Matching had strong effects on both behaviour and attitude measures. The effects of matching emotion versus mismatching emotion were so powerful that neither driver emotion nor voice emotion by itself had a consistent effect on drivers’ behaviour or attitude.

For driving behaviour matching is visible in the extraordinary reduction in accidents, half the accidents in matched cases (\( M = 3.39 \)) versus mismatched cases (\( M = 8.95 \)). Although there was a slight tendency for happy drivers to be better drivers, it failed to reach significance, \( M = 4.2 \) for Happy drivers and \( M = 5.9 \) for Sad drivers. There were no differences in accident rate based on voice emotion, \( M = 5.8 \) for the Happy voice and \( M = 6.5 \) for the Sad voice, see Table 2-7.

The effect on attention to the driving task (responses to horn honks) show a similar trend, drivers in matched cases responded almost twice as quickly than mismatched cases, see Table 2-7. It is not surprising to note that drivers without the in-vehicle system (\( M = 2.0 \)) responded quicker than drivers with an in-vehicle system (\( M = 5.3 \)). This confirms the literature (McCarley, Vais et al. 2001; Barón 2006) that speech systems in cars disrupts attention to the driving task, even if less than a visually based system (Lunenfeld 1989). There was no difference in attention to driving task based on driver emotion, \( M = 3.5 \) for Happy drivers and \( M = 3.7 \) for Sad drivers, \( F(1,54) = .14, p < .71 \). Neither did the data show any significant difference in attention to driving task based on voice emotion, \( M = 4.4 \) for Happy voice and \( M = 7.4 \) for Sad voice, \( F(1,36) = .07, p < .79 \).
Drivers that were paired with a matched voice in the in-vehicles system communicated much more with the system than drivers paired with a mismatched voice. It is interesting to note that the voice of the in-vehicle system said exactly the same thing in all conditions. This suggests that drivers who interacted with an in-vehicle system with a voice similar to themselves or their own emotional state were able to speak more while avoiding accidents. Driver emotion alone (M = 4.5 for Happy drivers and M = 4.5 for Sad drivers) or voice emotion alone (M = 4.5 for Happy voice and M = 4.5 for Sad voice) did not influence drivers willingness to communicate.

For the attitude measures, the data reveal the same trend as for the behavioural measures. Matching driver emotion with voice emotion has a positive effect on drivers’ attitude, whereas driver emotion and voice emotion alone had very little effect. Drivers in matched cases perceived themselves as more attentive drivers and that the in-vehicle system made them more positive and confident in matched cases than in mismatched cases. There was no effect on the perception of being an attentive driver based on driver emotion, M = 15.4 for Happy drivers and M = 14.5 for Sad drivers, or based on voice emotion (M = 14.9 for Happy voice and M = 15.0 for Sad voice). The data showed no influence of driver emotion, M = 59.7 for Happy drivers and M = 50.9 for Sad drivers, or voice emotion, M = 56.8 for Happy drivers and M = 53.8 for Sad drivers, on the assessment of the confidence of the in-vehicle system.

Matching the emotion of the driver with the emotion of the in-vehicle voice has an enormous effect on perception of quality. The measures of quality are the only measures where driver emotion and voice emotion have significance on their own. To assess the effect size of the measures, I use Partial $\eta^2$. Assessment of the quality of the in-vehicle system show that the matched effect is strong with a Partial $\eta^2$ of .61, whereas driver emotion had a Partial $\eta^2$ of .38, and voice emotion showed a Partial $\eta^2$ of .41. When assessing the quality of the car, the effect in the matched cases is once again stronger with a Partial $\eta^2$ of .48, than the effect to the voice emotion, Partial $\eta^2$ of .41, or the effect of the driver emotion with a Partial $\eta^2$ of .25.

The driving behaviour measures indicate that using the wrong or mismatched voice for the in-vehicle system seemed to distract drivers. They probably divided their attention and cognitive power between driving and unconsciously trying to figure out why someone interacting with them had such a different emotional state. Possibly the drivers also pondered why the voice on the in-vehicle system did not follow the social protocol and adapt to their (the driver’s) emotional state. The result was that drivers with mismatched in-car voice had more accidents and paid less attention to the driving task. This unease or distraction is also visible in the attitudinal measures.
where drivers in matched cases felt more at ease, liked and rated the quality of both in-vehicle system and car higher than drivers in mismatched cases.

The results from this experiment consistently show that finding the appropriate in-vehicle voice for the driver’s emotion stood out as the most critical factor in enabling a safe and engaging driving experience. Much more so than the drivers’ initial emotion or the quality of the in-vehicle voice alone. This suggests that people find it easier and more natural to attend to voice emotions that are consistent with their own (for a similar discovery in persuasion research, see (Wegener, Petty et al. 1995)). Emotion-inconsistent voices are arguably more difficult to process and attend to than emotion-consistent ones; in other words, interacting with an inconsistent voice takes more cognitive effort (Gong 2000). It then follows that listening to or conversing with Chris the Virtual Passenger was more distracting in the mismatched case, leading to poorer driving performance, less attention to the road, and a less comfortable experience.

It would be interesting to see what the effects of an emotionally neutral voice would be in comparison to an emotionally matched voice? Would the effect be between the negative effect of a mismatched voice and the positive effect of a matched voice? Would an emotionally neutral voice be seen as disinterested and non-engaging with a negative effect? Or would the emotionally neutral voice be seen as calm and soothing with a positive effect?

The data clearly show that there are positive effects of positive emotional states. A happy driver drives a little better than an unhappy driver. This must be seen as a challenge and an invitation to designers of cars and in-vehicle systems to design systems that makes drivers happy. Results presented here show that changing the voice to match the driver for valence (positive – negative emotion) with a neutral arousal level has a positive effect in a short 15 – 20 minute driving scenario. It is doubtful that matching a voice to the driver on both valence and arousal has the same positive effect – an angry voice for an angry driver? Results presented in the previous study show that familiarity and possibly loss of anonymity works for angry drivers. It is clear that more studies need to be conducted to answer questions on if and how to use voices and speech based in-vehicle systems to promote safer driving.

2.4 Making Driving and In-Vehicle Interfaces Safer and Better

Auto manufacturers spend a great deal of time and money attempting to make cars safer. They install anti-lock brakes, adaptive steering and gas pedals, air bags, and various warning indicators. All of these design decisions require very long lead times to re-tool factories and often involve dramatic increases in cost.
The two studies described here demonstrate that some very simple, inexpensive and fully controllable aspect of a speech-based in-vehicle interface can have a dramatic influence on driver safety. Something as small as changing the paralinguistic characteristics of a voice in the in-vehicle system is sufficient to significantly improve driving performance. Even with the same words spoken at the same times, either by the same voice or by different voices, driver performance can be radically altered. The first study showed that by selecting a voice for the in-vehicle system that is known by the driver, there is an enormous positive impact on both driver attitude and driving performance. The second study showed that by simply changing the in-vehicle system’s voice to match the drivers’ emotional state the same effect – improved attitude and driving performance – can be achieved. Designers of in-vehicle systems have the potential to influence the number of accidents, the drivers’ perceived attention to the road, and the driver’s engagement with the car simply by changing either the voice or simply the tone of voice.

A key finding here is that the same voice cannot be effective for all drivers. For both actual and perceived performance – one voice does not fit all. The first study clearly shows that a familiar voice works better than an unfamiliar voice. The second study show that upset drivers clearly benefited from a sad and subdued voice, while happy drivers clearly benefited from a happy and energetic voice. This suggests two things: a) that voices in cars should be known by their drivers and b) that voices in cars must adapt to their drivers. To realize this, two important questions need to be addressed: 1) How can an interface detect driver emotion? and 2) How can that information be used most effectively?

2.4.1 Detecting Driver’s Emotion

How should cars assess the emotion of drivers? The basic emotions (fear, anger, sadness, joy, and disgust) are the most distinguishable, and therefore measurable, emotional states (both in emotion recognition systems as well as in post-interaction evaluations). Further, the basic emotions are less likely to vary significantly from culture to culture, facilitating the accurate translation and generalisability of emotion detection. Emotion-recognition technology provides a number of possibilities for determining the emotional state of drivers and other users of interactive systems. Sensing technologies include cameras to detect facial expression (Ekman and Friesen 1975) (e.g., road rage or other emotions), skin conductance and blood pressure detectors attached to the steering wheel or mouse (particularly good for arousal) (Picard 1997), technologies to measure head-nodding or other movement (e.g., driver drowsiness or general relaxation), and eye-tracking equipment (e.g., detects pupil expansion, which is associated with positive feelings (Hess 1972)). In certain contexts, prediction also offers another possibility for detecting drivers’ emotions. For
example, if a car’s safety system recognizes a potentially dangerous situation, the in-vehicle system can assume that the driver will soon be afraid or stressed and quickly adapt its voice characteristics (without any direct measurement of emotion). In less time-critical situations, such prediction could also be used in conjunction with direct measurement to improve emotion recognition accuracy (Picard 1997).

Voice analysis is particularly useful when there are in-vehicle systems or passengers to whom the driver speaks. Voice analysis systems can use the same set of vocal cues to detect emotion that people use to detect emotion such as fundamental frequency, speech-rate, pitch range, rhythm, and amplitude or duration changes (Scherer 1989; Ball and Breese 2000; Cowie, Douglas-Cowie et al. 2001). In my work with Christian Jones, our research considers in-car affective computing where the car can recognize driver emotion and respond intelligently to support the driving experience such as offering empathy or providing useful information and conversation. The emotion recognition (ER) system (Jones and Jonsson 2005; Jones and Jonsson 2005; Jones and Jonsson 2005; Jones and Jonsson 2007; Jones and Jonsson 2008) we work with uses 10 acoustic features including pitch, volume, rate of speech and other spectral coefficients. The system then maps these features to emotions such as boredom, sadness, grief, frustration, extreme anger, happiness and surprise, using statistical and neutral network classifiers. The emotion recognition system uses changes in acoustic features representative of emotional state whilst suppressing what is said and by whom. It is therefore speaker independent and utterance independent and can be readily adapted to other languages.

The performance of the emotion recognition system was tested (Jones and Jonsson 2007; Jones and Jonsson 2008) using a conversational in-vehicle system. Even though engine noise, brake screech, indicators, sirens etc was mingled with the input to the emotion recognition system, the system detected the selected range of emotions; boredom, sadness, anger, happiness and surprise. However for most of the drive the emotional readout was neutral/natural. When challenged in the drive by obstacles in the road, other drivers, difficult road conditions and pedestrians, strong emotions (aroused emotional states) were observed from the drivers. The performance of the automatic emotion recognition system was assessed in two second windows by comparing the output from the automatic acoustic emotion recognition system with assessments from trained experts in recognizing affective cues in speech. The results show 65% match, 15% equivalent (sad, happy) and 20% mismatch between emotion recognition system and human expert. Given the short speech windows and the requirement of inertia in the system, not to be perceived as psychotic by shifting between happy and angry in two second intervals, this technology shows good promise. The results can most certainly be improved by reducing the number of
recognized emotional states from five in this case, to say three, thereby reducing the marginal of error.

The critical issue using automatic emotion detection is accuracy, and as my own and others results show, drivers are sensitive to accuracy (see chapter 3). Where even small amounts of inaccuracies impacts – in particular male drivers – trust and reliance on an in-vehicle system. While focusing on multiple indicators leads to better accuracy than assessments of only a single indicator of emotion, even the most complete, wide-ranging, and intelligent systems are not as good as the average person (Picard 1997). Thus, any reactions to the driver’s emotion must be tempered by the knowledge that the detection system’s judgments are not perfect.

2.4.2 Changes of Voice Characteristics
With knowledge of driver emotion the in-vehicle system can modify its response both in the words it uses but also the presentation of the message by stressing particular words in the message and speaking in an appropriate emotional state. Once the detection system has made a determination of the driver’s emotion, the next step for the designer is to decide how the in-vehicle system should respond to that emotion. One useful strategy is to change the voice of the system in step with the driver, that is, to exhibit empathy. Empathy greatly fosters relationship development, as it communicates support, caring, and concern for the welfare of another. A voice which expresses happiness in situations where the driver is happy and sounds subdued or sad in situations where the driver is upset would strongly increase the connection between the driver and the in-vehicle voice system (and hence the car). Although rapid response to emotion (or predicted emotion) of the driver can be effective, there are a number of dangers in this approach. In the human brain and body, emotion can change in seconds (Picard 1997). If someone tells a joke to a sad person, he or she may become momentarily happy but will fall back into their sad state relatively quickly. Conversely, happy drivers may become frustrated as they must slam the brakes after expecting to zoom through a yellow light, but their emotion may quickly switch back to feeling positively. If the voice in the car adapted immediately and forcefully to the driver’s emotions, drivers would experience such bizarre occurrences as the voice of the in-vehicle system potentially changing its characteristics in mid-sentence. People or voices that manifest emotions or change the way of speaking at that rate would dramatically increase cognitive load (Mullennix, Bihon et al. 2002). It would also constantly activate new emotions in the driver and most likely be perceived as psychotic. While this can be entertaining when performed by manic comedians like Robin Williams or Jim Carrey, it is psychologically exhausting and disturbing when encountered in daily life. An in-vehicle system with these properties
and behaviour would immediately be marked as manic-depressive and hard to interact with as opposed to empathetic!

To make the voice of an in-vehicle system an effective interaction partner, the designer must take longer term emotions, such as anger and frustration, into account. These long term emotions (or moods) tend to bias feelings, cognitive strategies, and processing over a longer time (Davidson 1994). Long term emotions also tend to be more stable and can even serve as a filter through which both internal and external events are appraised. A person in a good mood tends to view everything in a positive light, while a person in a bad mood does the opposite. So, when assessing driver’s responses to a voice, designers should consider the biasing effects of the long term emotions. Drivers entering a car or interacting with a voice-based product in a good mood, for instance, are more likely to experience positive emotion during an interaction than drivers in a bad mood. Thus, emotion in technology-based voices must balance responsiveness and inertia by orienting to both momentary and longer term emotions.

Changing the system voice to a familiar voice when a driver is angry and upset would leverage the same relationship building as for the matching emotions. It would, however, also be possible to simply always use a voice that is familiar for each individual user of an in-vehicle system – this could be the same or different voices when a car has multiple drivers. In this case the system leverages its ability to know its driver rather than measure the anger and frustration of the driver by analyzing utterances.

### 2.5 Stabilizing the Driver

People are who they are: Therefore grounding interface design in stable characteristics of individuals (called “personas” (Cooper 1999)) can simplify the design process and allow for standardization of interfaces. At the same time, humans are what they feel: Emotion, influenced by every encounter with the physical, psychological, and social world, plays a critical role in determining peoples’ performance, knowledge, beliefs, and feelings. Every voice, synthetic or real, emanating from people or attached to interfaces, creates and interacts with an emotion-driven individual who will think, feel, and succeed (or fail) based on the sensitivity of the design to the user’s affective reality.

Using a familiar voice and matching the voice of the in-vehicle system to the drivers’ emotions had enormous consequences. Using a familiar voice in an in-vehicle system had a positive effect on driving behaviour and attitude. The familiar voice stabilized the drivers, made drivers less angry and frustrated than drivers with the unfamiliar
Driver Emotion and Properties of Voices

voice. Interacting with the familiar voice also seemed less distracting than interacting with an unfamiliar voice. This resulted in better driving performance with fewer accidents and more attention to traffic regulations. Furthermore, drivers perceived the quality of both in-vehicle system and the car to be higher with the familiar voice. The overall effect is that a familiar voice in an in-vehicle system has the ability to have a positive influence on drivers’ trust and liking as well as improve driving safety. Providing there is a realistic way of assigning a familiar voice to each driver or group of drivers, this is also a low cost alternative to promote safer driving.

Drivers who interacted with a voice matched to their own emotional state (energetic voice for happy drivers and subdued voice for upset drivers) had less than half as many accidents on average as drivers who interacted with mismatched voices. The effects of matching an in-car voice to the driver emotion were more powerful than driver emotion alone. Although there was a slight tendency for happy drivers to be better drivers, even this effect was unimportant compared to the effects of matching. In other words, finding the appropriate in-car voice for the driver’s emotion stood out as the most critical factor in stabilizing the driver and hence enabling a safe and engaging driving experience. Providing there is a reliable way of determining the driver’s emotional state, this is a cost effective means to safer driving.

The common result from both experiments presented here, is that no single voice – or tone of voice – fits all, and that selecting the voice will depend on the properties of the driver. It becomes clear that the car needs to know its driver. This can be accomplished in multiple way such as initializing the in-vehicle system when the car is purchased, monitoring the driver, updating the system through interactions, having the driver initialize the system with their personalized module each time the car is started, or some combination hereof. The important factor here is to determine, once the system is initialized, is exactly what and how much the system should do to stabilize the driver. Should we rely on speech based in-vehicle systems - are all drivers comfortable interacting with an in-vehicle system? How should familiar voices be selected - is it possible to find a familiar voice for each individual driver? Should in-vehicle systems change the car environment – climate, music, amplitude – as well as conversation strategy in response to a driver’s state? Should in-vehicle systems change interaction strategy based on passengers on the car? How quickly – and how often - should the system change to adapt to the driver? These and many more questions need to be answered before an in-vehicle system can truly be useful in stabilizing the driver.
2.6 Acknowledgements and Publications

For the first study, I selected and designed the inducement materials. Helen Harris (Toyota InfoTechnology Center US) and I selected and/or adapted the pre- and post-driving questionnaires, and I selected the emotional self-report questionnaires (DES). I designed the driving course and the speech prompts. I organized and analyzed all data from the study. The study was setup and run at Oxford Brookes University. Mary Zajicek (Oxford Brookes) and Helen Harris (Toyota InfoTechnology Center US) worked with me to setup the experiment and to schedule and run all subjects.

For the second study, I selected and designed the inducement materials, and I selected the attention to driving task. I selected the questionnaires for emotional self-reporting (DES), and I was one of many hands in the design of pre and post driving questionnaires. I worked with a group of students to schedule and run subjects (at Stanford University). The data was entered and organized by students. I analyzed the data together with Clifford Nass.

For the work on a framework for generating and indexing emotional voice data, I worked on inducement and inducement verification strategies. Database setup and indexing was mainly done by Aaron Masters.

For the work on automatic emotion recognition, I have mainly been involved in selecting emotions to focus on, collecting data for the system, and designing responses to affective input. All work on the automatic recognition system was done by Christian Jones.

Material from these studies has been published as follows:


Voice Emotion. CHI '05 extended abstracts on Human factors in computing 
systems, ACM Press, New York NY.

Jonsson, I. M. and F. Chen (2006). Detecting and Responding to Emotional Speech in 
Cars. 16th World Congress on Ergonomics International Ergonomics 
Association, Maastricht, Netherlands.

Generating and Indexing Induced Emotional Voice Data. International 
Conference on Language Resources and Evaluation, Genoa, Italy.

Recognition for In-Car Conversational Interfaces. Universal Access in 
Human-Computer Interaction, Beijing, China.

emotions in older car drivers. Affective and Emotion in Human Computer 
Interaction C. Peter and R. Beale: 229 -240.
3 Accuracy of Information in Vehicles

3.1 Driving, Trust and Quality of Information

Driving requires focused attention and timely decision making for appropriate manoeuvres and relies to some extent on well-timed and accurate information. When designing an in-vehicle information system it is important to ensure that the information given to the driver does not negatively affect cognitive processing and driving performance, and that the information is presented in a way that does not distract attention from the driving task.

There are different types of in-vehicle information system, ranging from conversational, to interactive to purely informational. These in-vehicle systems come with their own critical design considerations since there are fundamental differences between an in-vehicle computer that simply generates information and a system that engage drivers into conversation. While the critical problem of conversational and interactive voice systems tends to be issues of understanding the driver, i.e., speech recognition and contextual interpretation of the utterance, the critical problem in informational speech output systems is their accuracy and reliability (Kantowitz, Hanowski et al. 1998; Tseng and Fogg 1999). This is particularly important when the information focuses on current and near-term events, such as road conditions, traffic patterns, and suggested routes.

There is an extensive literature on trust in automation, where most are in agreement that accuracy of information is crucial in building and retaining trust. Proposed benefits of trust include better task performance (Golembiewski and McConkie 1975) and the ability to cooperate (Deutsch 1962; Argyle 1991). There are a number of taxonomies for trust in human-machine relationships (Barber 1983; Rempel, Holmes et al. 1985; Muir 1987; Muir 1994). These taxonomies tend to distinguish three categories of trust (Lee and Moray 1994): (1) trust that is based on observed consistency of behaviour (persistence or predictability); (2) trust that is based on a belief in competence (competence or dependability); and (3) trust that is based on faith in purpose or obligation (fiduciary responsibility or faith). While faith-based trust is difficult to establish for computers, these taxonomies suggest that trust by consistency and competence can be achieved by explaining the purpose, capability, and reliability of the software or the system to the user.
Experiments by Muir (1987) and Lee and Moray (1994) investigates the effects of errors on ratings of trust and willingness to use automation. The results show correlations between ratings of trust and reliance on automation. Bonsall and Perry (1991) have shown that the quality of information determines acceptance of that information, where accuracy is just one property of quality. Lee and Moray (1994) also examined the relationship between trust, self-confidence and manual control abilities in operating a simulated semi-automatic plant. They showed that trust combined with self-confidence predicted the operators' strategy, and that automation is used when trust exceeds self-confidence and manual control when the opposite is true. This emphasizes the importance of building trust in automated or semi-automated systems based on the information provided by the system. It is also important that the information is presented in a manner that is credible and builds trust. Aspects of this will be studied in later chapters, whereas in this chapter effects of the quality of information will be investigated. Important issues in the quality of information are for instance, how quality affects attention and how quality affects trust. This becomes interesting to study in a domain (traffic and driving) where ephemeral information potentially has a great impact and makes up a large percentage of the domain information.

Kantowitz et al. (1998) showed that varying the accuracy of the information in an in-vehicle information systems influences driving performance. The study varied the quality of advice for two levels only, 100 percent versus 77 percent accuracy, and the system provided two types of information, harmless and potentially harmful. Potentially harmful information was related to the driving task, such as road and traffic condition, whereas harmless information was not. They found that tolerance for inaccurate harmless information was much higher than the tolerance of inaccuracies in harmful information.

To investigate how driving related accurate and inaccurate information affects drivers performance as well as attitude, we designed a speech-based in-vehicle system that with levels of accuracy ranging from 65% to 100%. This expands the work of Kantowitz et al. (1998) by adding more granularity in the accuracy of the system, and uses measures to investigate behavioural and attitudinal responses from the driver. The information provided by the system was purely driving related, what Kantowitz termed “potentially harmful” information, but it can definitely also be seen as potentially helpful information. The system was a pure information system, designed to give information that pertained to road conditions and traffic situations. There was no attempt to engage the driver in a dialogue, and there was no need for the driver to initiate communication or respond to the system’s comments and suggestions.
In particular we designed to driving scenario and the in-vehicle system to answer the following questions:

- Will drivers notice inaccuracies in an in-vehicle information and warning system?
- Will the accuracy level of information from an in-vehicle information and warning system affect the driver’s performance? Are there threshold values?
- Will an in-vehicle information and warning system improve driving performances over driving without the system?
- Will the in-vehicle information and warning system affect the driver’s attitude towards the car or the system itself?
- Will the voice of the system affect drivers?

Based on previous results by Kantowitz, Lee and Moray, we would predict that varying the accuracy of information presented by an in-vehicle information and warning system would affect both the driver’s attitude about the trustworthiness and expertise of the system, and the driver’s performance.

### 3.2 In-vehicle Information and Hazard Warning System

The in-vehicle information and hazard warning system was implemented on a driving simulator. A specially-designed driving scenario, 60,000 feet (18.3 km) long, was built using a storyboard. The driving scenario took the drivers along the countryside on road that varied between four and two lanes. The four lane road had two lanes in each direction and the two lane road had one lane in each direction. Small and larger villages were located along the roadside, some had intersections with stop signs and others had intersections with traffic-lights.

The road varied between two lanes in each direction (as depicted to the left) and one lane in each direction. This stretch of road was country side driving, with no houses.

Depicted below in Figure 3-2 is a typical intersection with moderate oncoming traffic.
and cross traffic. The right side of Figure 3-2 show a section of the driving scenario with high-rises, shops and pedestrians.

Figure 3-2: Intersection and small city

Most villages were small with only stop signs for intersections; larger villages had intersections with traffic lights, sometimes with multiple intersections and traffic lights. Speed limits ranged from 20 to 60 miles per hour (32 – 97 kmh), and density of pedestrians, parked cars and traffic varied from none to crowded. The storyboard placed hazardous road conditions and complex traffic events, such as intersections, foggy and windy stretches, road work and school zones in the driving scenario.

Figure 3-3 depicts a potentially hazardous traffic situation. There is roadwork spanning the two middle lanes of the road. This leaves a narrow space for cars to pass. Drivers were forced to reduce speed and to merge with the traffic in the rightmost lane to clear the roadwork area.

Figure 3-3: Road work

There were also obstacles such as cars parked half-way off the road, and bicyclists placed at scripted locations along the road. All these obstacles and traffic events were spaced to prevent information and warning prompts to overlay a previous situation.
Figure 3-4 shows a vehicle that pulls out right in front of the drivers. To avoid an accident, the driver needs to reduce speed and/or change lanes. The picture to the right in Figure 3-4 depicts a situation where the driver is stuck behind a couple of slow trucks on a narrow and windy road. The driver needs to slow down and find appropriate locations to overtake the slow vehicles.

An in-vehicle information system was created based on the storyboarded driving scenario. All information prompts were of an ephemeral type that could feasibly be supplied by police reports or weather reports, together with more permanent information such as the location of school zones and speed limits. The prompts were all worded to inform and warn about activities, features and hazards related to the road or the driving environment (Jonsson, Nass et al. 2004). The prompts were inserted at points in the driving scenario at which potential road hazards or traffic events were about to be noticed.

The voice prompts for the in-vehicle information system were recorded by a female speech talent that spoke in a neutral tone of voice. A female voice was selected because it is typically used in car system, partially because the combination of greater variance in voice frequency (Childers and Wu 1991) and the lessened interference between car noise and female voices makes female voices easier to hear in a car than male voices (Nixon, Anderson et al. 1998; Nixon, Morris et al. 1998). Female voices are also found to portray a greater range of urgency when used in a warning system (Edworthy, Hellier et al. 2003).

The information content of the speech based information and warning prompts used by the system, ranged from informing the driver of the current speed limit, to warning the driver about potential road hazards. Special care was taken to phrase the prompts according to level of urgency as exemplified by the prompts listed below. For a full list of prompts see Appendix B.
Accuracy of Information in Vehicles

“Current speed limit is 60 miles per hour”
“Warning, there is a fallen tree in the middle of the road”
“Caution, there is heavy fog up ahead”.
“Caution, watch for merging traffic”
“Caution, you are entering a school zone”
“You are approaching a construction zone”
“Warning, there is a car stopped in your lane”

To investigate if there was a threshold of errors that would be acceptable, there were four (five including a system without prompts) variations of accuracy of the in-vehicle information system. These variations were created by introducing two types of errors into a 100% correct in-vehicle information system:

- **Omissions**—warning prompts were removed from hazardous situations and traffic events.
- **Random information**—warning prompts were randomly placed in locations lacking hazards or traffic events (e.g., there would be a warning “You are approaching a construction zone” on a stretch of road without road work).

<table>
<thead>
<tr>
<th>Accuracy Level</th>
<th>Correct Prompts</th>
<th>Omitted Prompts</th>
<th>Randomly Placed Prompts</th>
</tr>
</thead>
<tbody>
<tr>
<td>100 % correct in-vehicle system</td>
<td>33</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>88 % correct in-vehicle system</td>
<td>29</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>76 % correct in-vehicle system</td>
<td>25</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>64 % correct in-vehicle system</td>
<td>21</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>No in-vehicle system</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Please note that the same prompts were used in all variations of the in-vehicle information system. When a warning prompt was omitted, it was later randomly placed in the scenario. Thus by moving one prompt, two errors are created, the first by omitting to warn the driver of a road hazard, and the second by warning the driver when there is no hazard.

The first variation of the system was 100% accurate: every prompt was relevant to current events in the driving simulator. For example, when the driver was approaching a foggy section, the system would produce the prompt, “Caution, there is heavy fog ahead”. The second to fourth variations of the in-vehicle system were designed by
moving information prompts and hence deliberately adding errors. All systems used the same recorded prompts as the 100% accurate system, but some of them were moved to points at which they would not make sense. By moving the prompts according to Table 3-1, driving courses with in-vehicle information systems were created where the correctness of the information given by the speech prompts were 88%, 76% and 64% respectively. The fifth variation of the in-vehicle system was silent, and did not give any information and warning prompts to the driver during the driving session in the simulator. This variation was added to give a baseline for driving behaviour. This enabled comparisons to be made with regards to improvements or degradation of driving performance when using the in-vehicle warning system.

### 3.3 Design of Study

#### 3.3.1 Overview

The experiment was designed as a 5 (in-vehicle information system: 100% accurate, 88% accurate, 76% accurate, 64% accurate, and Silent Condition) x 2 (participant gender, male and female), between-participant design. Gender is included as a factor in this study. The reason for this is that previous studies (Jonsson, Nass et al. 2004; Nass, Jonsson et al. 2005) where reaction to voices from in-vehicle systems were studied, show significant gender differences in driving behaviour for drivers aged 18-25. Male drivers tended to drive faster, have more accidents and disobey traffic regulations more than female drivers. These studies are summarized in chapter 5, where the observed gender differences are also discussed.

All participants were informed that the study would last approximately an hour and that the experiment was set-up as two driving simulator sessions. Participants answered an online questionnaire after each session. The entire study was conducted in one room where both the driving simulator and the computer with the online questionnaire were located.

#### 3.3.2 Participants

There were a total of 100 participants, 50 female and 50 male, recruited to participate in the study. All of the participants were Stanford students and in the age group 18 – 25. They all had a driver’s license and between one and five years of driving experience. All participants gave informed consent and were debriefed after the experiment.
3.3.3 Experimental Apparatus

STISIM Drive from Systems Technology Inc. was used as the driving simulator. Drivers were seated in a real car seat and ‘drove’ using a Microsoft Sidewinder forced feedback steering wheel and pedals consisting of accelerator and brake. The simulated journey was running on a computer and back-lit-projected onto a six-foot diagonal rear-projection screen. The same simulator set-up was used for both driving sessions.

The driving simulator was instrumented to collect and save data on driving behaviour. The simulator saved data on different indicators of driving performance such as obedience to traffic laws (e.g., stopping at stop signs, not excessively exceeding the speed limit), accidents, brake behaviour and lane deviation. All verbal feedback from the participants while in the driving simulator and self-reported data on the assessment and acceptance of information and warning prompts was collected. Attitudinal data was gathered by a set of both standard (Rubin, Palmgreen et al. 1994) and newly-designed questionnaires, see Appendix B. The questionnaires were constructed to collect information on source credibility of the in-vehicle information system, trust, reliability and quality of the system and the effects of the in-vehicle system on the quality of the car.

3.3.4 Procedure

Each participant started the experimental session by signing the consent form. They were then seated at the driving simulator and drove a training course with verbal guidance from the experimenter. The training driving course was 5 000 feet (1.5 km) and took approximately three minutes to finish. The purpose of the short introductory driving session was to familiarize participants with the control and feedback from the driving simulator. This course also provides screening for participants that suffer from simulator sickness (Bertin, Guillot et al. 2004). None of the 100 participants felt nauseous or discomfort during or after the training course, and continued to fill in the first questionnaire with general information such as gender and age in addition to driving experience.

All participants used the same simulator configuration and the same driving course in the experiment. The course was 60 000 feet (18.3 km) long and took on average 25 minute to complete. Participants were randomly assigned to one of five conditions. Four had in-vehicle information and warning system: 1) 100% accurate, 2) 88% accurate, 3) 76% accurate, and 4) 64% accurate. All of these versions used the same prompts; accuracy was manipulated simply by changing the location of the prompts (as noted in Table 3-1). The fifth system did not have an in-vehicle system.
After the main driving session, participants filled in two web-based online questionnaires. The first questionnaire asked about the quality of both the car and the information system, and the second asked about the effects of driving the car with the in-vehicle information system. For the fifth variant of the in-vehicle information system, i.e. when the participant was driving without any information and hazard prompts, the questions referring to the information and hazard prompts of the information system were omitted. Participants used a desktop computer to complete and submit the web-based questionnaires. All participants were debriefed at the end of the experimental session.

3.4 Measures

3.4.1 Video Game Experience
Using a driving simulator as the experimental setting, it is of interest to control for use on video games. Participants were asked if they played videogames, which type videogames and how many hours they played per day.

3.4.2 Prior Driving Experience
Information about participants’ real life driving experience was collected as part of the first questionnaire. Participants were asked to self report how many years they had been driving, the number of miles they typically drive per week, the number of accidents they had had, and the number of tickets they had received.

3.4.3 Driving Performance
The simulator, STISIM Drive, automatically generated and gathered a large variety of data on driving behaviour. We collected two sets of driving behaviour measures. First a set of driving behaviours often used in driving simulator studies including drivers’ lane deviation behaviour, the use of the brakes, speed and time to first accident. The second set of driving behaviour data was incident related, matching what drivers are judged by in real traffic including accidents and adherence to traffic regulations. For this study, we focused on the second group of measures since these are most suitable for the low fidelity of the driving simulator that was used. Thus, three measures for the most dangerous behaviours: number of collisions, number of off-road accidents, and obedience to two important traffic laws; adherence to traffic lights and adherence to stop signs were monitored in the study. These measures were uncorrelated, see below. However, we also report on averages for the first group of measures, lane deviation, brake usage and speed as well as time to first accident.

All three of these serious negative behaviours are more common in a driving simulator than in actual driving; indeed, we would be startled to find even two
occurrences of these behaviours if the experiment involved driving in real traffic. One key reason for dramatically higher rates of dangerous behaviours in this simulator study is that we created an extremely difficult driving course to study variance in driving performance: A simple driving course of the same length would fail to generate any variance in poor driving behaviour.

3.4.4 Quality of the In-vehicle Information System and the Car

The quality-rating of the in-vehicle information system came from an index of terms that were rated using 10-point Likert scales (1 = Describes Very Poorly to 10 = Describes Very Well). Participants were asked “How well do the following words describe how you feel about the voice system?” The index included the following terms: High in quality, I would use it, I would buy it, I would like to have it, I would recommend it, and Useful. The index was very reliable (Cronbach’s α = .95). The rating for the car used the same terms and Likert scales, and was based on the question, “How well do the following words describe how you feel about the car?” The index was very reliable (Cronbach’s α = .89).

3.4.5 Credibility and Trust in the In-vehicle Information System

Perceived Accuracy of In-vehicle Information System: Participants were asked to estimate the accuracy of the in-vehicle information system. The measure was based on the question, “Please indicate what percentage of time the system was accurate” in a scale from 0 to 100 in intervals of 10.

Source Credibility and Trust: In general, one can distinguish trust and credibility even though these two concepts are linked. When a source is considered trustworthy, or better when an individual trusts a particular source, it is more probable that this individual will accept pieces of knowledge (beliefs) coming from that source. Trustworthiness is a property of a source while credibility should be considered a property of a piece of information. When a piece of information is credible, it is convincing and believable.

In attempting to distinguish these two concepts, the questionnaire that were used included credibility items developed byBerlo and McCroskey (Rubin, Palmgreen et al. 1994), as well as items from the Individualized Trust Scale (Rubin, Palmgreen et al. 1994). All of these scales are based on pairs of adjectives which anchor seven-point Likert scales from 1 - 7. The pairs of adjectives are antonyms.

We created three measures of credibility and trust: a source credibility measure and a trust measure based on standard scales, and a combination credibility-trust. The third
measure was a based on a factor-analysis of the items in the standard scales, where most of these items of credibility and trust loaded on the same factor.

Source credibility was made up of the concepts of Safety, Dynamism and Qualification by Berlo (Rubin, Palmgreen et al. 1994) and the concepts of Authoritative and Character by McCroskey (Rubin, Palmgreen et al. 1994). The index was very reliable, (Cronbach’s α = .87). Trust was made up of the items in the Individualized Trust Scale (Rubin, Palmgreen et al. 1994), and the index was very reliable (Cronbach’s α = .89).

The third measure was made up of items from credibility and trust that loaded on the same factor were: reliable/unreliable, informed/uninformed, qualified/unqualified, intelligent/unintelligent, valuable/not valuable, expert/inexpert, honest/dishonest, trustworthy/untrustworthy, “trustful of this system/not trustful of the system”, safe/unsafe, candid/not candid, not deceitful/deceitful, straightforward/not straightforward, considerate/inconsiderate, honest/dishonest, reliable/unreliable, faithful/unfaithful and careful/careless. The index was very reliable (Cronbach’s α = .82).

3.5 Results

The data from the study were analyzed using the accuracy of the driving system, gender, and their interaction as independent variables. In regression analysis, the significance of the interaction effect is assessed after the main effects via increment to R² (Jaccard and Turrisi 2003). To investigate threshold values, part of the data was analyzed using two-way ANOVA’s with gender and in—vehicle system as between participants’ factors. The data was analyzed in two ways: with accuracy and gender as factors, and for driving with or without system and gender as factors.

Analysis of the effect of the in-vehicle information system on driving performance was divided into two parts. First we examined the effect of an in-vehicle system relative to no system at all. Then we looked at the effects of the level of accuracy of the in-vehicle system on driving performance. Reported in the results are significance levels of .05, these area marked by “*” in the tables. Some trends (non significant results) are also reported, these are marked by “+” in the tables. Note that M is short for Mean and SD is short for Standard Deviation.

3.5.1 Video Game Experience

To assess the potential influence of video games on the experience with the driving simulator, we include the number of hours spent playing video games (see Table 3-2). There was no significant difference across conditions, $F(4, 90) = .31, p < .87$. 
Accuracy of Information in Vehicles

Table 3-2: Means and Standard Deviations - Video Games

<table>
<thead>
<tr>
<th>Conditions</th>
<th>Hours of Video games per day</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
</tr>
<tr>
<td>100%</td>
<td>.40</td>
<td>.68</td>
</tr>
<tr>
<td>88%</td>
<td>.35</td>
<td>.60</td>
</tr>
<tr>
<td>76%</td>
<td>.30</td>
<td>.73</td>
</tr>
<tr>
<td>64%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Silent</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

3.5.2 Prior Driving Experience

To ensure that there were no initial differences between the driving abilities of participants across conditions, we compared years of driving experience, average number of miles driven per week, number of prior accidents, and the number of tickets across conditions (see Table 3-3). The data show no significant differences for years of driving experience, $F(4,90) = .50, p < .74$, for miles driven per week, $F(4,90) = 1.6, p < .19$, for accidents, $F(4,90) = .67, p < .61$ or for the number of tickets, $F(4,90) = .61, p < .65$.

Table 3-3: Means and Standard Deviations - Real Life Driving Experience

<table>
<thead>
<tr>
<th>Conditions</th>
<th>Years of Driving Experience</th>
<th>Average Miles Driven per Week</th>
<th>Number of Accidents</th>
<th>Number of Tickets</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
<td>SD</td>
</tr>
<tr>
<td>100%</td>
<td>4.3</td>
<td>1.7</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>88%</td>
<td>3.9</td>
<td>1.5</td>
<td>77.50</td>
<td>38</td>
</tr>
<tr>
<td>76%</td>
<td>4.6</td>
<td>1.3</td>
<td>100</td>
<td>63</td>
</tr>
<tr>
<td>64%</td>
<td>4.4</td>
<td>1.7</td>
<td>135</td>
<td>118</td>
</tr>
<tr>
<td>Silent</td>
<td>4.0</td>
<td>2.1</td>
<td>127</td>
<td>95</td>
</tr>
</tbody>
</table>

3.5.3 First Incident, Lane Deviation, Speed and Brake Usage

The regression analysis of the first group of driving performance data (Time to first incident speed, lane behaviour and brake usage) show very few effects.

**Time to first incident:** There were no significant effects of accuracy, gender or when driving with or without a system on time to first incident, ($\beta = .08, p < .41, \beta = -.14, p < .18$, and $\beta = .08, p < .40$). See Table 3-4 for means and Table 3-5 for $\beta$-Coefficients. Neither were there any interaction effects of gender and existence of system on time to first incident $\beta = -.24, p < .39$, or any interaction effects of gender and accuracy or system time to first incident, $\beta = -.27, p < .26$. See Table 3-4 for means and Table 3-5 for $\beta$-Coefficients.
Social and Emotional Characteristics of Speech-based In-Vehicle Information Systems: Impact on Attitude and Driving Behaviour

Table 3-4: Means, Standard Deviations - Lane Deviation and Time to First Accident

<table>
<thead>
<tr>
<th>Conditions</th>
<th>Average lane deviation (ft)</th>
<th>Time to first accident (ms)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
</tr>
<tr>
<td>100%</td>
<td>10.1</td>
<td>1.7</td>
</tr>
<tr>
<td>88%</td>
<td>10.3</td>
<td>1.6</td>
</tr>
<tr>
<td>76%</td>
<td>10.0</td>
<td>1.4</td>
</tr>
<tr>
<td>64%</td>
<td>9.8</td>
<td>1.9</td>
</tr>
<tr>
<td>Silent</td>
<td>9.7</td>
<td>1.0</td>
</tr>
</tbody>
</table>

**Lane Deviation:** The data show no effects of accuracy, gender, or when driving with or without the in-vehicle warning system on lane behaviour (Accuracy: $\beta = .13, p < .21$, Gender: $\beta = .008, p < .94$, and System: $\beta = -.112, p < .27$). See Table 3-4 for means and Table 3-5 for $\beta$-Coefficients.

There were no interaction effects of gender and accuracy or system on lane behaviour, $\beta = -.26, p < .27$. Neither were there any interaction effects of gender and existence of system on lane behaviour, $\beta = -.20, p < .38$. See Table 3-4 for means and Table 3-5 for $\beta$-Coefficients.

Table 3-5: $\beta$ Coefficients - Lane Deviation and Time to First Accident

<table>
<thead>
<tr>
<th>Independent Variables</th>
<th>Lane Deviation</th>
<th>Time to first accident</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy</td>
<td>.13</td>
<td>.08</td>
</tr>
<tr>
<td>Gender $R^2$</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>.008</td>
<td>-.14</td>
</tr>
<tr>
<td></td>
<td>.02</td>
<td>.03</td>
</tr>
<tr>
<td>Accuracy * Gender</td>
<td>-.26</td>
<td>-.27</td>
</tr>
<tr>
<td>Increment to $R^2$</td>
<td>.03</td>
<td>.04</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Independent Variables</th>
<th>Lane Deviation</th>
<th>Time to first accident</th>
</tr>
</thead>
<tbody>
<tr>
<td>System vs. no system</td>
<td>.112</td>
<td>.08</td>
</tr>
<tr>
<td>Gender $R^2$</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>.008</td>
<td>-.14</td>
</tr>
<tr>
<td></td>
<td>.013</td>
<td>.03</td>
</tr>
<tr>
<td>System/No System * Gender $R^2$</td>
<td>-.202</td>
<td>-.24</td>
</tr>
<tr>
<td>Increment to $R^2$</td>
<td>.02</td>
<td>.04</td>
</tr>
</tbody>
</table>

**Speed:** The results for average speed show main effects of gender and also interaction effects of gender with accuracy of system and gender with having a system or not. See Table 3-6 for means and standard deviation.

The data for speed show a main effect for gender, female drivers consistently drive slower than male drivers, ($\beta = .34, p < .001$), see Table 3-7. There is no main effect of accuracy of in-vehicle system and no interaction.
There is also an interaction trend of gender and of driving with an in-vehicle system. This is only a trend; it fails to reach significance, and should be further investigated. The trend indicates that male drivers drive slower with a system than without a system ($\beta = -.25, p < .08$), see Table 3-7, and probably reflects a floor effect since women generally drives slower than men.

### Table 3-6: Means and Standard Deviations - Speed and Brake Behaviours

<table>
<thead>
<tr>
<th>Conditions</th>
<th>Average Speed (mph)</th>
<th>Average Brake usage ($0 = $no usage and low values=heavy usage)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
</tr>
<tr>
<td>100%</td>
<td>Male</td>
<td>34.1</td>
</tr>
<tr>
<td></td>
<td>Female</td>
<td>31.2</td>
</tr>
<tr>
<td>88%</td>
<td>Male</td>
<td>32.7</td>
</tr>
<tr>
<td></td>
<td>Female</td>
<td>30.3</td>
</tr>
<tr>
<td>76%</td>
<td>Male</td>
<td>31.5</td>
</tr>
<tr>
<td></td>
<td>Female</td>
<td>30.4</td>
</tr>
<tr>
<td>64%</td>
<td>Male</td>
<td>33.8</td>
</tr>
<tr>
<td></td>
<td>Female</td>
<td>30.3</td>
</tr>
<tr>
<td>Silent</td>
<td>Male</td>
<td>36.2</td>
</tr>
<tr>
<td></td>
<td>Female</td>
<td>29.4</td>
</tr>
<tr>
<td>All Conditions</td>
<td>Male</td>
<td>33.7</td>
</tr>
<tr>
<td>All Conditions</td>
<td>Female</td>
<td>30.3</td>
</tr>
</tbody>
</table>

### Brake Usage: The results for average brake usage show main effects of gender and also interaction effects of gender with accuracy of system and gender with having a system or not. See Table 3-6 for means and standard deviation.

The data for usage of brakes show main effects of gender and interaction effects of gender and driving with a system, as well as gender and accuracy.

Female drivers on average brake less than male drivers ($\beta = -.24, p < .015$), see Table 3-7. The interaction effects show that female drivers brake more with a higher accuracy system, ($\beta = -.39, p < .01$), and that they brake more when driving with a system than when driving without a system ($\beta = .31, p < .03$), see Table 3-7.

The data was also analyzed using two-way ANOVA’s – treating the accuracy levels as discreet groups instead of a continuous variable- to identify threshold values. Results from the first set of driving performance measures show no significant effects for lane deviation and time to first accident. The ANOVA’s show no threshold values...
for when accuracy starts influencing speed, but indicate a threshold value of accuracy on brake behaviour.

**Table 3-7: β Coefficients - Speed and Brake Behaviours**

<table>
<thead>
<tr>
<th>Independent Variables</th>
<th>Speed average</th>
<th>Brake Usage</th>
<th>Independent Variables</th>
<th>Speed average</th>
<th>Brake Usage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy</td>
<td>-.07</td>
<td>-.05</td>
<td>System vs. No System</td>
<td>-.08</td>
<td>-.04</td>
</tr>
<tr>
<td>Gender</td>
<td>.34*</td>
<td>-.24*</td>
<td>Gender</td>
<td>.34*</td>
<td>-.24*</td>
</tr>
<tr>
<td>R²</td>
<td>.12*</td>
<td>.08*</td>
<td>R²</td>
<td>.12*</td>
<td>.08*</td>
</tr>
<tr>
<td>Accuracy * Gender</td>
<td>-.40</td>
<td>.59*</td>
<td>System/No System * Gender</td>
<td>-.35+</td>
<td>.53*</td>
</tr>
<tr>
<td>Increment to R²</td>
<td>.14*</td>
<td>.12*</td>
<td>Increment to R²</td>
<td>.14*</td>
<td>.12*</td>
</tr>
</tbody>
</table>

The data for time to first accident and lane deviation show no significant effects of accuracy ($F(4,90) = 1.2, p < .30$ and $F(4,90) = .55, p < .70$ respectively), no gender effects ($F(1,90)=1.9, p < .18$ and $F(1,90) = .006, p < .95$ respectively). There was no significant effect of driving with or without a system on time to first accident and lane deviation ($F(1,96) = .71, p < .40$, and $F(1,96) = 1.2, p < .27$), see Table 3-4 for means and standard deviation.

The ANOVA’s show effects on brake behaviour when driving with or without a system. There is a clear gender effect $F(1,96) = 10.5, p < .002$, and an interaction effect with gender and driving with a system $F(1,96) = 4.2, p < .04$, see Table 3-6. Female drivers brake more than male drivers, and female drivers with a system brake more than female drivers without a system. The data show no effect of having a system or not on brake behaviour, $F(1,96) = .21, p < .65$.

The ANOVA’s show no influence of accuracy on brake behaviour, $F(4,90) = .18, p < .95$. There is however a clear gender difference for brake behaviour $F(1,90) = 6.6, p < .01$, and an interaction effect gender and brake behaviour $F(4,90) = 3.6, p < .009$ (See Table 3-6). Female drivers brake more than male drivers, and female drivers also brake more with a system with higher accuracy than with a system with lower accuracy. When analyzed as discreet groups based on accuracy levels, the data show a clear change in female drivers brake behaviour somewhere between 64% accuracy and 76% accuracy. This threshold change in brake behaviour cannot be seen in male drivers.
The ANOVA for speed show no effect of having a system or not on speed ($F(1,96)=.76, p < .39$), a clear gender effect ($F(1,96) = 16.1, p < .001$). There is also an interaction trend $F(1,96) = 3.5, p < .06$, that needs further investigation, see Table 3-6. This interaction trend indicates that male drivers drive slower with a system than without a system. There is no effect of accuracy on speed ($F(4,90) = .52, p < .72$), there is however a clear gender effect ($F(1,90)=12.5, p < .001$) (male driver driving faster then female drivers) with no interaction effects.

Summary: there are no main effects or interaction effects on lane behaviour and time to first incident. The data show that male drivers drive faster than female drivers, and the female drivers brake more than male drivers. Female drivers are also more sensitive to accuracy and existence of a system so that they brake more with a system and with higher accuracy (threshold between 64% and 76% accuracy). There is an interesting trend that notes that male drivers drive slower with a system than without a system. This needs to be further investigated.

### 3.5.4 Collisions, Accidents and Obedience of Traffic Regulations

The regression analysis of the second set of driving performance measures (Collisions, accidents and obedience to traffic rules) reveal an interesting set of main effects and interaction effects. For clarification, a collision is defined by the driving simulator as an accident with an object (obstacles, vehicles etc) that happens on the road. An off-road accident is defined by the driving simulator as an accident with an object (vehicle, house, obstacle etc) that happens outside of the paved road area.

The overall number of collisions, off-road accidents and disobedience of traffic rules are most often higher in driving simulator studies than in real life. For this study, the data show that the overall average number of collisions per driver was 2.5 (SD=1.6), with the number of collisions ranging from 0 to 7 (See Table 3-8).

#### Table 3-8: Means and Standard Deviations - Driving Performance Measures

<table>
<thead>
<tr>
<th>Independent Variables</th>
<th>Collisions Mean</th>
<th>Collisions SD</th>
<th>Off-road Accidents Mean</th>
<th>Off-road Accidents SD</th>
<th>Disobedience to traffic rules Mean</th>
<th>Disobedience to traffic rules SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>100%</td>
<td>2.0</td>
<td>1.3</td>
<td>0.8</td>
<td>0.9</td>
<td>0.7</td>
<td>0.7</td>
</tr>
<tr>
<td>88%</td>
<td>2.2</td>
<td>1.6</td>
<td>0.7</td>
<td>0.7</td>
<td>0.6</td>
<td>0.7</td>
</tr>
<tr>
<td>76%</td>
<td>2.4</td>
<td>1.3</td>
<td>0.8</td>
<td>0.9</td>
<td>0.6</td>
<td>0.6</td>
</tr>
<tr>
<td>64%</td>
<td>2.9</td>
<td>1.1</td>
<td>0.9</td>
<td>0.8</td>
<td>0.7</td>
<td>0.7</td>
</tr>
<tr>
<td>Silent</td>
<td>2.9</td>
<td>1.8</td>
<td>1.3</td>
<td>1.1</td>
<td>0.5</td>
<td>0.6</td>
</tr>
<tr>
<td>TOTAL</td>
<td>2.5</td>
<td>1.6</td>
<td>0.89</td>
<td>0.93</td>
<td>0.63</td>
<td>0.73</td>
</tr>
</tbody>
</table>
There were an average of 0.89 off-road accidents per driver (SD =0.93), with the number of off-road accidents ranging from 0 to 4 (See Table 3-8). Finally there was an average of 0.63 incidents where drivers disobeyed traffic laws (SD=0.73), with the number of violations ranging from 0 to 3 (See Table 3-8).

Results are reported first as a comparison of driving with a system or without a system, then the effect that the level of accuracy had on driving behaviours.

Comparing system and no system: Incorporating an in-vehicle information and warning system into the car compared to having no system at all did not have a direct effect on collisions, ($\beta = -0.43$, $p < .17$) see Table 3-9. There was no effect for gender ($\beta = .01$, $p < .90$). However, there was a marginally significant interaction, such that females benefited from having an in-vehicle system ($\beta = -0.32$, $p < .03$), while there was no effect for males ($\beta = .04$, $p > .77$) see Table 3-9. There was a main effect on off-road accidents of having an in-vehicle system in the car ($\beta = -0.22$, $P < .03$), see Table 3-10. This led to fewer off-road accidents than having no system at all. There was no effect for gender ($\beta = .08$, $p < .45$) and no interaction, see Table 3-9.

There were no main effects on adherence to traffic rules when driving with a system ($\beta = .10$, $p < .30$) or for gender ($\beta = .06$, $p < .52$) and no interaction ($\beta = -0.02$, $p < .94$), see Table 3-9.

Table 3-9: $\beta$ Coefficients for In-vehicle System vs. No System – Driving performance

<table>
<thead>
<tr>
<th>Independent Variables</th>
<th>Collisions</th>
<th>Off-Road Accidents</th>
<th>Disobedience to Traffic Rules</th>
</tr>
</thead>
<tbody>
<tr>
<td>System vs. No System</td>
<td>-0.14</td>
<td>-0.22*</td>
<td>0.10</td>
</tr>
<tr>
<td>Gender R*</td>
<td>0.01</td>
<td>0.08</td>
<td>0.06</td>
</tr>
<tr>
<td></td>
<td>0.02</td>
<td>0.06</td>
<td>0.02</td>
</tr>
<tr>
<td>System/No System * Gender Increment to R2</td>
<td>0.45*+</td>
<td>0.22</td>
<td>-0.02</td>
</tr>
<tr>
<td></td>
<td>0.03*+</td>
<td>0.01</td>
<td>0.00</td>
</tr>
</tbody>
</table>

Effect of level of Accuracy: Greater accuracy of the in-vehicle information and warning system led to significantly fewer collisions ($\beta = -0.23$, $p < .04$), see Table 3-10. There was no effect for gender ($\beta = .11$, $p < .30$), but there was an interaction effect, ($\beta = -1.9$, $p < .03$) (see Table 3-10). Female drivers were pretty unaffected by the level of accuracy ($\beta = 0.02$, $p < .92$), while male drivers benefited from higher
levels of accuracy ($\beta = -.47, p < .002$). The data suggests that accurately alerting drivers to hazards in the road can help them to avoid these hazards.

Table 3-10: $\beta$ Coefficients for Accuracy of In-vehicle System – Driving performance

<table>
<thead>
<tr>
<th>Independent Variable</th>
<th>Collisions</th>
<th>Off-Road Accidents</th>
<th>Disobedience to Traffic Rules</th>
</tr>
</thead>
<tbody>
<tr>
<td>System Accuracy</td>
<td>-0.23*</td>
<td>-0.06</td>
<td>-0.10</td>
</tr>
<tr>
<td>Gender</td>
<td>0.11</td>
<td>0.14</td>
<td>0.06</td>
</tr>
<tr>
<td>R^2</td>
<td>0.06</td>
<td>0.02</td>
<td>0.01</td>
</tr>
<tr>
<td>System Accuracy * Gender</td>
<td>-1.90*</td>
<td>1.92*</td>
<td>1.82*</td>
</tr>
<tr>
<td>Increment to R2</td>
<td>0.05*</td>
<td>0.06*</td>
<td>0.06*</td>
</tr>
</tbody>
</table>

Accuracy of the system did not generally lead to fewer off-road accidents ($\beta = -0.06, p < .59$), nor did gender have a direct effect ($\beta = .14, p < .23$), see Table 3-10. However, there was a significant interaction between accuracy and gender ($\beta = 1.9, p < .03$). Specifically, female drivers benefit from increased accuracy ($\beta =-0.34, p<.03$), while there was no effect for male drivers ($\beta =-0.18, p<.28$), see Table 3-10.

There was no main effect for accuracy of an in-vehicle information and warning system on how drivers adhere to key traffic laws ($\beta = -.10, p < .40$). There was no main effect for gender ($\beta = .06, p < .61$), see Table 3-10. However, there was a significant interaction such that female drivers were increasingly better at driving legally with increasing levels of accuracy of the system ($\beta =-0.31, p<.05$), while there was no significant effect for male drivers ($\beta =.16, p>.34$).

Analyzing the data for the second group of measures (Collisions off-road accidents and obedience to traffic rules) using ANOVA’s enabled identification of threshold values. Male drivers have fewer collisions in accuracy levels 88% and 100% than in all other conditions, for female drivers the threshold value is below 64% (if it exists). The data show that female drivers benefit when driving with a system (regardless of accuracy level) as opposed to driving without a system. The lowest accuracy level for any of the systems was 64%, and female drivers have fewer collisions even with this poor accuracy than when driving without a system.

There is a clear positive effect of driving with a system than driving without a system on off-road accidents ($F(1,96) = 5.0, p < .03$). All drivers had fewer off road accidents driving with the system than without the system. This would indicate that the threshold is low, equal to or lower than 64%, since driving with any system is better than driving without the system. There is no main effect of accuracy on off road accidents ($F(3,72) = .30, p < .82$).
For adherence to traffic regulations, there were no main effects of existence of system ($F(1,96) = 1.1, p < .30$) accuracy ($F(3,72)= .51, p < .68$), or gender ($F(1,90)= .41, p < .52$). Neither were there any interaction trends or effects.

Summary: the pure existence of an in-vehicle system - regardless of accuracy – resulted in fewer off-road accidents for all drivers.

Higher accuracy of the system resulted in fewer collisions for all drivers, fewer off-road accidents for female drivers, and better adherence to traffic regulations for female drivers.

The data show potential thresholds of accuracy where drivers benefit from the system. Male drivers have fewer accidents (collisions and off-road accidents) for accuracy levels of 88% and higher. The accuracy threshold for female drivers is much lower, in this study - with a lowest accuracy level of 64% - the mere existence of a system seems to benefit female drivers; they experienced fewer accidents and showed better adherence to traffic regulations for even a 64% accurate system.

### 3.5.5 Quality of the In-Vehicle System and Car

Source Credibility, Trust and Quality of the Car: The data showed no effects on perceived car quality for driving with an in-vehicle system versus driving without the system ($\beta = -.15, p < .14$), neither was there an effect of gender ($\beta = -.03, p < .75$) or any interaction, see Table 3-11.

<table>
<thead>
<tr>
<th>Independent Variables</th>
<th>System Quality</th>
<th>Car Quality</th>
</tr>
</thead>
<tbody>
<tr>
<td>System vs. No System</td>
<td>-</td>
<td>-0.15</td>
</tr>
<tr>
<td>Gender</td>
<td>-</td>
<td>-0.03</td>
</tr>
<tr>
<td>R*</td>
<td>-</td>
<td>0.02</td>
</tr>
<tr>
<td>System/No System * G</td>
<td>-</td>
<td>-0.01</td>
</tr>
<tr>
<td>Gender</td>
<td>-</td>
<td>0.00</td>
</tr>
</tbody>
</table>

Accuracy, however, had a dramatic effect on credibility/trust. Greater accuracy was clearly associated with greater credibility ($\beta = .55, p < .001$), greater trust ($\beta = .45, p < .001$), and with greater combined credibility-trust ($\beta = .48, p < .001$) in the in-vehicle information and warning system (see Table 3-12).

There was no effect for gender for any of these measures (Credibility - $\beta = -.07, p < .50$, Trust - $\beta = -.07, p < .50$, and Combined Credibility Trust - $\beta = -.12, p < .30$ respectively) and no interactions (see Table 3-12).
Higher accuracy led to a higher quality rating of the in-vehicle system ($\beta = .35, p < .002$), and higher accuracy led to more positive perceptions of the car as well ($\beta = .23, p < .05$), see Table 3-12. There was no gender effect for the quality measures $\beta = .01, p < 1.0$, and $\beta = -.05, p < .70$ respectively, and no interactions.

There is a significant correlation between the perception of the quality of the in-vehicle system and the quality of the car ($r = .35, p < .001$). The “halo effect” of extending positive feelings from the system to the car has been noted in previous studies of speech based in-vehicle information systems (Jonsson, Nass et al. 2004; Nass, Jonsson et al. 2005) and suggests that the decision about which in-vehicle system to use is an important one for car manufacturers.

**Perceived level of Accuracy:** Drivers’ assessment of the accuracy of the information system was significantly lower than the actual values for 100% and 88% accurate systems.

There were no significant differences at the lower levels of accuracy of the system, see Table 3-13. Drivers could quite accurately assess the percentage of errors in information presented by the system for the more inaccurate systems.

There were no gender differences in these assessments, see Table 3-13. Both male and female drivers assessed the accuracy of the system similarly and without noticeable or significant differences for all actual accuracy level.
Table 3-13: Perceived and Actual Accuracy of In-vehicle Information System

<table>
<thead>
<tr>
<th>Actual Value of Accuracy</th>
<th>Perceived Value of Accuracy</th>
<th>Difference between actual and perceived Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean %</td>
<td>SD</td>
</tr>
<tr>
<td>Male</td>
<td>90</td>
<td>6.7</td>
</tr>
<tr>
<td>Female</td>
<td>90</td>
<td>9.4</td>
</tr>
<tr>
<td>Total 100</td>
<td>90</td>
<td>8.0</td>
</tr>
<tr>
<td>Male</td>
<td>77</td>
<td>16</td>
</tr>
<tr>
<td>Female</td>
<td>77</td>
<td>20</td>
</tr>
<tr>
<td>Total 88</td>
<td>77</td>
<td>17.5</td>
</tr>
<tr>
<td>Male</td>
<td>72</td>
<td>10</td>
</tr>
<tr>
<td>Female</td>
<td>76</td>
<td>9.7</td>
</tr>
<tr>
<td>Total 76</td>
<td>74</td>
<td>9.9</td>
</tr>
<tr>
<td>Male</td>
<td>67</td>
<td>10</td>
</tr>
<tr>
<td>Female</td>
<td>66</td>
<td>12</td>
</tr>
<tr>
<td>Total 64</td>
<td>66</td>
<td>10.8</td>
</tr>
</tbody>
</table>

3.6 Discussion

The data show clear evidence that the accuracy of an in-vehicle information and warning system has effects on driving performance. Accuracy also has significant effects on attitudes toward the system and the car.

Perhaps the most striking result in this research is the dramatic differences between females and males with respect to the effects of the accuracy of the in-vehicle system on their driving behaviour. For female drivers, greater accuracy improved all three measures of driving performance: collisions, off-road accidents, and obedience to traffic rules. Even for low levels of information accuracy, the mere presence of the in-vehicle system helped female’s driving behaviour with respect to reducing collisions and off-road accidents. The situation for male drivers was very different. The accuracy of the in-vehicle system had a positive effect on the number of collisions and the presence of an in-vehicle system had a positive effect on off-road accidents, but there were no other effects. Thus, while accurate in-vehicle systems can be of benefit for all drivers, they are clearly helpful for females even when their accuracy is limited.

It is interesting to note that even though the number of off-road accidents is affected by the existence of an in-vehicle system, collisions are also affected by high-accuracy
systems. Off-road accidents are often defined as single accidents, where a driver swerves off the road, very seldom are other vehicles involved. Collisions, however, are accidents that happened on the road and almost always involve other vehicles. This type accident is often defined by the driver accidently ending up in the lane of opposing traffic.

There were no significant differences in lane behaviour and very few variations in speed and brake behaviour. The most noteworthy effect for female drivers was that they tended to use their brakes more with increased accuracy of in-vehicle system and when driving with an in-vehicle system rather than driving without the system. Providing that the female drivers braked to avoid prompted obstacles, this brake usage can help explain the effect of the information systems accuracy level on driving performance. For male drivers, the data show a trend for male drivers driving slower when driving with an in-vehicle system than driving without a system. Speed is a driving parameter that is correlated with traffic incidents, so providing that male drivers slowed down as a result of listening to the system, this change in driving behaviour could help explain the improved driving performance.

The ANOVA’s for the driving behaviour measures also show some indications of threshold values for accuracy of information in the in-vehicle hazard and warning system. The gender differences are clear for the thresholds values as well. The data indicates a threshold value of around 90% for male drivers to see a significant reduction in the number of collisions, for female drivers, the threshold – if it exists and is not simply the mere existence of a system – lies around or lower than 64%. For off-road accidents, the threshold for reduction in accidents is low; all drivers showed improvements in the measure for accuracy levels of 64%. This would indicate that the threshold is equal to or lower than 64%. Female drivers also showed sensitivity to accuracy, having fewer off-road accidents with increased accuracy of the in-vehicle system. Finally, the data did not reveal any threshold values for adherence to traffic regulations; there was no effect of accuracy or the existence of a system. Based on the ANOVA’s the threshold values are different for female and male drivers, while the mere existence of an in-vehicle system - regardless of accuracy - seems beneficial for female drivers, male drivers show most improvement in driving performance when accuracy is 90% or more.

3.6.1 Gender Differences
Why should females and males be so differently affected by the in-vehicle system? One explanation that can be ruled out is that the gender differences are due to differences in perceived accuracy or credibility. All participants were sensitive to, and reported inaccuracies in the in-vehicle information system. The data show no gender
Social and Emotional Characteristics of Speech-based In-Vehicle Information Systems: Impact on Attitude and Driving Behaviour

Differences in drivers’ assessments of how accurate the systems were. Similarly, females were neither more forgiving nor more critical of the credibility of the in-vehicle system than males. The explanation for these results must thus be found elsewhere and there are two main theories that can explain why female drivers would be more sensitive to the statements of the system as well as its accuracy.

A series of studies, conducted in the US, by Meyers-Levy (1988; 1989) demonstrates that women feel more responsible for attending to information from other sources—they will listen even when the source is considered unreliable. In essence, women try to take all information into account and engage in a detailed elaboration of the content while men try to ignore it and focus on broad message themes. Women are ‘comprehensive processors’ who attempt to assimilate all available information before rendering judgment. Females’ tendency to scrutinize information would lead them to be more sensitive to accuracy differences than men. Similarly, females’ tendency to take all information into account would lead even a somewhat inaccurate in-vehicle system to be better than none at all.

Tannen’s (1990) extensive work on gender and communication provides a different explanation. Please note that Tannen’s studies, just as Meyers-Levy’s studies, were conducted in the US, and that the potential for culture differences on gender differences were noted in translated version of her work (Adelswärd 1991). According to Tannen, women use speech as a means of establishing relationships, while men use speech to establish status and a hierarchy of superiority. Men are more comfortable giving information and advice than accepting advice or information; women are equally comfortable giving and receiving information. Women tend to be inclined to do what is asked of them, while men tend to resist even the slightest hint that anyone, especially a woman, is telling them what to do. Under this view, the quality of the source would be less important than the fact that they were being offered unsolicited advice. Indeed, there would be a general tendency for males to think neither about the quality of the source nor the quality of the information: the in-vehicle information and warning system would simply seem to be adopting a position of hierarchical superiority as it does not even allow the driver to respond. Women, conversely, would welcome the information in that the system is clearly trying to be helpful.

An alternative third theory, also based on studies in the US, is related to the voice of the in-vehicle system. This theory predicts that since the voice of the in-vehicle system was female, males tends to be more dismissive of the system than did females (Nass, Moon et al. 1997; Nass and Brave 2005). Although both genders seem to have negative stereotypes about females when it comes to issues of driving (Nass and Brave 2005), females do exhibit social identification effects toward female voices,
i.e., a positive bias toward females, even when they come from machines (Nass and Brave 2005). If males dismiss the female voice, they would be less likely to be influenced by its accuracy and less likely to be influenced by even the voice’s presence.

The data from the presented research cannot determine which of these three explanations are applicable. However, one way to distinguish whether the second theory is correct is to compare the current system to a system that is initiated and controlled by the driver. Tannen’s theory would suggest that allowing male drivers to control the system by voice might allow the driver to move beyond mere concerns with hierarchy and to more closely assess the in-vehicle system’s statements. The first and third explanations suggest that voice input would not affect the discernment of male drivers.

The third theory can be evaluated relative to the first two by replicating the presented experiment using a male voice in the in-vehicle information and warning system. Although one would have to be concerned about the fact that male voices might seem unusual and thereby would elicit different responses, the effects of stereotyping should lead male drivers to pay much more attention to the voice. The other two theories would not predict differences between male and female voices. Furthermore, the study was done in the US, and the cultural differences between the US and the rest of the world can be quite significant. This needs to be taken into consideration when discussing how the results transfer and the potential for culture differences on gender differences. These cultural differences are not necessarily tied into a language; there are major cultural differences between same language countries such as the US, the UK, and Australia. However the study of gender and cultural differences are research topics in their own right, and a more elaborate discussion on the topic is outside the scope of this study.

Beyond the behavioural differences associated with the accuracy of the in-vehicle information system, there are also credibility and trust effects. The accuracy of the system was an extremely strong determinant of perceived credibility for both female and male drivers, even though drivers were not always aware of how accurate the system was. The data showed a spill-over – or contagion - effect from perception of in-vehicle system to perception of the car. This trend has been noted in all studies of speech-based in-vehicle system that I have conducted, and is a note to automobile manufacturers to think carefully about branding when introducing in-vehicle systems. Furthermore, there is also a risk that a new in-vehicle system could affect any existing in-vehicle systems in addition to the car. As an example, inaccuracies in a new in-vehicle hazard and warning system could negatively influence the trust of advice and
information other system in the car such as navigation systems or car-status information systems. The inaccuracies in one system have the potential to reduce the credibility of all systems as well as the car itself.

3.7 Conclusion and Further Questions

The results indicate that variations in the accuracy of an in-vehicle information system affect not only performance, but also the drivers’ attitude regarding the information system and the simulated car. The data also shows that all drivers listened to the information given and noted fairly accurately how often it was right or wrong. So the answer to the question “Will drivers notice inaccuracies in an in-vehicle information and warning system?” is “yes.”

The second question was whether the accuracy level of information from an in-vehicle information and warning system affect the driver’s performance. The data show that both female and male drivers were aware of the accuracy level of the in-vehicle information and warning system, and that the accuracy level of the system affected especially female drivers. Accuracy affected female drivers on three measures out of three, collisions, off-road accidents and obedience to traffic regulations. Male drivers were only affected on one measure, collisions.

Answering the third questions, Will an in-vehicle information and warning system improve driving performances over driving without the system?, the data show that female drivers did benefit from having the in-vehicle information and warning system in the car, and male drivers did not. Female drives benefited from having a system in the car on two out of three measures, collisions and off-road accidents. Male drivers only showed an improvement on one measure out three measures, off-road accidents. Because male drivers listened to the information from the system without taking advice, it was similar to not having the system in the car.

In answer to the fourth question, “Will the in-vehicle information and warning system affect the driver’s attitude towards the car or the system itself?, the results clearly show that trust and credibility decreased with increasing levels of error. The accuracy level of the in-vehicle information system influenced the perceived quality of both the in-vehicle system and the car for all drivers. The quality of the car and quality of the in-vehicle information system showed a correlation, previous studies (Jonsson, Nass et al. 2004; Nass, Jonsson et al. 2005) show the same trend: assessed quality of in-vehicle systems affect assessments of the car itself. An open question is whether strongly branding the system as distinct from the car would mitigate this effect. Another area to investigate is that even though clear gender differences were found for driving performance; where female drivers were less sensitive to inaccuracies than
male drivers - these differences were not visible in the trust and credibility of the system or the perceived quality of the car or in-vehicle system. In these assessments, female and male drivers agreed.

There are limitations to the described experiment that might affect the interpretation of the results. One limitation of the current research is that a driving simulator rather than an actual car was used for the experiment. As noted earlier, the study found many more poor driving behaviors than would be found in a typical driving situation. This is based on a having drivers drive through an extremely hard driving scenario, where hazards and traffic events are more numerous than most drivers experience (if ever) over longer periods of time. We believe that the results are nonetheless relevant because results obtained in driving simulator studies are indicative of likely patterns in actual cars since they are due to inattention and lack of judgment. Driving simulator studies are furthermore often used as the first screening of various car assistance systems and information systems since it is less costly than testing in a real car. It allows for iterations of development cycles, and as seen from this study, there are many more parameters that can be fine-tuned before testing the system in real cars. We cannot extrapolate to the actual levels of poor driving in real life, but we are confident that the patterns associated with accuracy and presence or absence of the in-vehicle systems do extrapolate to less difficult driving situations. Furthermore, it would be impossible to do this study with actual cars: There is no current system that can accurately or reliably judge road conditions at either 100% or even at any prescribed level.

Another limitation of this research is that the in-vehicle system was limited in its functionality – it presented only one type of information. It is important to determine if people’s perceptions of in-vehicle road warning systems would be affected if the system provided other types of information that might be much more accurate than the road assessment system—e.g., navigation or problems with the car—or much less accurate—e.g., a restaurant or music recommendation system. It would also follow that it is important to determine if the use of different voices for different type information affects perception and performance. This study does not cover the long term effects of using an in-vehicle hazard and warning system. Attitude and allowances might change over time with the usage of a system so that drivers can become either more or less tolerant to inaccuracies of information. This attitude change in connected in how we accept and interact with communication media ranging from TV to interactive software to computer supported cooperative work-tools. There might also surface a difference in attitude towards inaccuracies in different categories of information. Map based data such as road layout (lane-count, intersections, and exit-ramps/on-ramps) tends to be more static than incident based
data that is more ephemeral (accidents, road-work, and weather-patterns) and tolerance of errors might be higher for dynamic data than for static data.

Future work for perception of accuracy vs. inaccuracy of information should also investigate the timing for providing information and warnings to the driver of a vehicle. There could be interaction effects of improper timing and perceived inaccuracy of the information. Critical factors for timing include age of driver, visibility, road conditions, number of passengers in the car, and the type of warning. Are older adult drivers in need of more information then young drivers? Should information be phrased and timed differently for different categories of drivers? Should states and traits of drivers influence the choice and timing of information? Are stationary objects easier to react to than moving ones? Are vehicles in front less of a problem than those in the rear? These issues are crucial for driver safety and should be considered when testing and designing in-vehicle information systems that focus on assisting the driver with the driving task.

Most important, however, is to find the balance between technology and efficacy. Even if a system that is 50% accurate is a remarkable technological feat, if drivers are not positively influenced by it, it is a wasteful expense. Even worse, it is possible that very inaccurate systems will actually be worse than no system at all. Especially if the driver has adverse reactions to the faulty information or systems such that drivers make inappropriate manoeuvres listening to the system when there is no danger and hence create a hazardous situation, or if the driver simply starts to ignore all information, even information that is correct and helpful. Similarly, if a system is perceived as inaccurate or undesirable, regardless of its actual performance, drivers will be dissatisfied with both the system and the car. The bottom line is that even the technologically-best system may not satisfy or help all drivers: While an in-vehicle hazard and warning system represent exciting technological advances, its deployment should be guided by significant caution. The results from this study suggest that the accuracy has to reach a certain level; for male drivers in the US the data indicates high levels around 90%, while a level as low as 64% seems sufficient for female drivers. Even though the results from this study indicate gender differences and threshold values for accuracy in an in-vehicle hazard and warning system, the results might not transfer across cultures, or even across age groups. The drivers in this study were all Americans in the age span of 18-25 and sensitivity to inaccuracies of information could well be different for other age groups of drivers or for drivers from other cultures. Furthermore, an inaccurate system might have negative effects on the perceived value of other in-vehicle systems. There is a risk, based on previously documented spill-over effects, that drivers will transfer both positive and negative perceptions to the car and possibly also other systems in the car.
3.8 Acknowledgements and Publications

The study was designed by me and Helen Harris at Toyota InfoTechnology Center. I designed the driving courses and together, Helen and I designed the speech prompts. The questionnaires were selected by me and some of the pre-and post-driving questionnaires were adapted by me to match the study. The experiment was setup and run by me, Helen and a group of Stanford students. I am responsible for the ANOVA’s. Clifford Nass, Stanford University, did the initial regression analysis, and I continued to analyze the entire dataset using linear regression.

Material from this study has been published as follows:


4 Older Adult Drivers and Voices for In-Vehicle Systems

4.1 Older adults and Driving

The car is an integral part of everyday life for many people and Li and Smith (2002) emphasize that the car and safe driving practices provide the means for freedom of movement, independence, and mobility. The car is used to commute to work or other activities, and it is also in the US, the main form of transportation for grocery shopping and visits to medical institutions. In addition to using the car for necessities and errands, the car is also a tool that provides the means for socializing, leisure and also to lead an active outdoor life. Outdoor activities are most often located some distance away from the public transportation network. Furthermore, the car is often the preferred vehicle of transportation after physical activities such as hiking or skiing. To support older adults to uphold an active lifestyle, it is important to address issues in the design of cars and in-vehicle system that improve their driving performance and enables them to stay as safe drivers. The benefits of mobility for older adults are that they stay healthy, and hence autonomous, longer (Coughlin 2003).

The driving task places significant perceptual and cognitive demands on the driver, and the normal aging process negatively affects many of the perceptual, cognitive, and motor skills necessary for safe driving (Ponds, Brouwer et al. 1988; Brouwer 1993). As a group, older adults have fewer collisions than younger driver (Guerrier and Fu 2002). Taking into account miles driven, statistics show that drivers 75 and older are involved in twice as many collisions per mile than younger drivers (Retchin, Cox et al. 1988). Hakamies-Blomqvist, Raitanen and O’Neill (Hakamies-Blomqvist, Raitanen et al. 2002) however argued that the relationship between yearly mileage and number of accidents per mile is not linear. Drivers with large yearly mileages have fewer crashes per mile than drivers with low yearly mileages (Janke 1991). Dividing young and older drivers into groups based on yearly mileage, the age disadvantage disappears again, and older adult drivers instead tended to be safer drivers (Hakamies-Blomqvist, Raitanen et al. 2002).

There are indications suggesting that older adult drivers may have more difficulty in attending to the driving task, especially when required to make complex decisions (Rinalducci, Mouloua et al. 2003). Walker et al (Walker, Fain et al. 1997) showed that age affects decision-making and route selection and in particular that older adults need more time than younger drivers to make decision in traffic. This does not, however, affect the quality of the decision. Greater driving speed lowers the
confidence of the decisions for older drivers, and time pressure and stress exacerbate the problem. As long as there is sufficient time, no pressure, and task familiarity, older adults perform well even in complex traffic situations (Walker, Fain et al. 1997). Fast-paced and demanding traffic patterns are however stressful for older adults and merging provides an example where problems might occur. Older adults have difficulty seeing and determining the speed and distance to merging traffic (Guerrier, Manivannan et al. 1996; Cox, Broshek et al. 2002). Mobility and agility is also affected by normal aging and makes it harder for older adults to turn their necks enough to see merging traffic. These two problems, determine speed and distance, and agility, coupled with declines in attention and working memory makes it harder for older adults to acquire and maintain all the information needed for merging.

Elderly drivers might not experience problems while driving during daylight on a road with light traffic. However, driving at night on poorly marked roads in heavy traffic can be taxing. Older drivers also often feel both unsafe and insecure as drivers. This is to a large extent due to observations and concerns about declining abilities and in particular, vision (Johansson 1997). Owsley et al observed that older drivers with vision problems allocate more resources to focus their attention on the driving task; this often results in a failure to recognize visual information and a decline in driving performance (Owsley, Ball et al. 1991). This behaviour is most pronounced when the workload is high, and studies show that most accidents involving older drivers occur at intersections (Hakamies-Blomqvist 1993; Johansson 1997; Read, Ward et al. 2001). More recent studies with older adults in intersections also show that accuracy of speed and distance estimations are influenced by age as well as recency of accidents (Read, Ward et al. 2001). There are often multiple causes for accidents caused by older adults (Guerrier and Fu 2002) and a review of the literature on older adults, driving and accidents show that, in addition to intersections, common causes and dangerous risk factors include:

- Failure to maintain proper speed (Hakamies-Blomqvist, Siren et al. 2004)
- Backing and parking your car (Oxley and Mitchell 1995)
- Improper left turns (Knoblauch, Nitzburg et al. 1995)
- Merging and failure to yield right-of-way (Mitchell and Suen 1997)
- Confusion in heavy traffic, due to a combination of attention, decision making and reaction time (Ball, Owsley et al. 1998; Owsley, Ball et al. 1998)
- Inattention (Hakamies-Blomqvist 1994)
- Hesitation in responding to new traffic signs, signals, road markings or different traffic patterns and roadway designs (Guerrier, Manivannan et al. 1996; Cox, Broshek et al. 2002)
Older drivers are often aware of problems with driving and change their driving behavior accordingly, by refraining from driving in unfamiliar settings, driving in rush hour, driving in bad weather and driving at night. In a large study of 2311 older adults by Rabbitt et al. (Rabbitt, Carmichael et al. 1996) reported that most older adults ceased to drive due to health problems or lack of confidence.

It is interesting to note in this context that cars bought by older adults have more extras than cars that young people buy (Oxley and Mitchell 1995). In a Swedish study (Viborg 1999), similar results were found. When asked about their views about several in-vehicle information systems, older adults (65+) were more positive about using them than drivers in the age span of 30-45. The most popular systems were automatic speed adjustment, automatic distance keeping, help crossing intersections, left turn (right turn) assistance. It is interesting to note that older adult drivers were more willing to relinquish control to the vehicles, at least partly, since the two last systems take part control over the vehicle (Davidse 2006).

4.2 Programs to Support Older Adult Drivers

The growing number of older adult drivers in conjunction with the concern of road safety organizations has led to programs whose purpose is to facilitate driving for older adults. This includes the use of larger letters on street signs, improved lane markings, and offset left turns. Subsequent studies verified the efficiency of the larger letters and improved lane markings, but failed to show improvements of offset left turns (Guerrier and Fu 2002). There are also initiatives by the Institute of Consumer Ergonomics and Petzäll (Petzäll 1991) for better vehicle design to support older adults. Some of their suggestions include re-thinking designs for entry and exit to the car, moving handles, adjusting height of door, and swivel seats. These designs, in addition to adjustable equipment such as mirrors, seating, and steering wheel, might enable older adults to drive more safely and longer. In addition to infrastructure measures and car design, there is also the potential of Intelligent Transport Systems (ITS) to provide support for older adult drivers. ITS offers a broad range of communication based control, information and electronics technologies to help monitor traffic, provide navigation help for drivers, and otherwise assists drivers. Mitchell & Suen (Mitchell and Suen 1997) found five systems that might provide assistance with limitations in motion perception, peripheral and selection attention, Entemann & Kuting (Entemann and Kuting 2000) added a sixth system that might help with information processing and decision making. These systems are: 1) Collision Warning Systems – particularly for intersections, 2) Automated Lane Change and Merging Systems, 3) Blind Spot and Obstacle Detection, 4) Display traffic Signs in Vehicle 5) Intelligent Cruise Control, and 6) Driver Assist in Intersections.
Collision warning systems could potentially be helpful in intersection to focus the older adult driver attention to oncoming traffic. This could provide assistance with turns at intersections that cross oncoming traffic. Driving simulator studies show promising results when the timing of cross traffic turns can be individually adjusted (Oxley and Mitchell 1995). Systems that assist drivers to change lanes and merge are still under development and are not expected to be available until after 2010 (Mitchell and Suen 1997; Davide 2006). Examples of these systems are lane-change collision warning systems (LCCW) that alert drivers to the objects in the drivers’ blind spot. More advanced lane-change collision warning and avoidance (LCCWA) systems also automatically steers away (Regan, Oxley et al. 2001). Some of these systems, such as LCCW, are already on the market but there is no data to indicate how well older adult drivers are helped by them.

Systems that assist drivers with blind spot and obstacle detection are most appropriate for backing up or parking. When tested, these systems were well liked by older adults (Oxley and Mitchell 1995). To assist older adults with traffic signs displayed in the car would require investments both in standards and infrastructure (Mitchell and Suen 1997). Results from a study show that these systems tend to draw the drivers focus away from the roadway (Lee 1997). Combining cruise control with automatic distance keeping might support older adult driver. This system can either alert the driver or take control to keep distance. Data from a study (Brook-Carter, Parkes et al. 2002) show that even though older adult drivers used the system well, it did not lower their driving workload and hence may not be relevant for safe driving by older adults. The system proposed by Entmann & Kuting (Entemann and Kuting 2000) is a combination of a navigation system and a traffic information system. In addition to route information it will also provide timely information on traffic events with the goal to aid drivers to focus attention on the most important traffic event. Results from a study of the system show that the extra information was useful for drivers; it made their driving task easier and increased safety compared to a standard navigation system (Entemann and Kuting 2000). Even though most of these systems have had no to little user testing, Oxley & Mitchell (Oxley and Mitchell 1995) report that older adult drivers that participated in the Elderly and Disabled Driver Information Telematics project (EDDIT), found the systems useful and would pay to have the systems in their cars. This was confirmed by driving simulator studies by De Waard, Van der Hulst & Brookhuis (De Waard, Van der Hulst et al. 1999).

As with all new technology, even if the aim is to assist, there are critical design issues related to variation in driver preferences and abilities across countries, ages, and genders. To attract new users, technology must in addition to be seen as useful also be designed for ease-of-use, acceptance and trust. This is especially important when
targeting older adults, where new technology must be both accessible and affordable in addition to being acceptable. Furthermore, there is also a risk that even if a technology shows great promise it might could induce over-reliance and distract drivers from the driving task (Coughlin 2003).

In chapter 3, I describe how a hazard and warning in-vehicle information systems resulted in improved driving performance for young drivers (age group 18 – 25) when tested in a driving simulator (Jonsson, Nass et al. 2005; Jonsson, Harris et al. 2008). The hazard and warning system used pre-recorded speech prompts to inform and warn drivers. These speech prompts were recorded in a female voice and informed drivers about road hazards, road conditions and traffic events. The results from the study showed a clear benefit for younger drivers when driving with the system. Driving performance was greatly improved, measured by fewer accidents and better adherence to traffic regulations such as stopping for signs and traffic lights. In addition to better driving performance, the young drivers also liked and trusted the system.

For older adults, voice prompts and speech messages provide reminders concerning previous interaction for those with poor memories in the domain of computing (Zajicek and Hall 2000). Voice messages can play an important part in helping older people to execute everyday tasks or to help them execute them more efficiently. The messages can provide useful information about the environment and current events, which older people may not readily absorb for themselves. Speech support is also used to help with strategizing, making contextually relevant suggestions, i.e. what to do next in computer interactions, and to provide warnings in safety critical situations. This type of speech support could easily be visualized as part of an in-vehicle information system to help older adult with the driving task.

The positive effect of an in-vehicle hazard and warning system on young drivers, provided an indication that such a system would work for other age groups as well. Unknown factors here are of course the efficacy of speech messages in cars for older adults, the selection of voice for the hazard and warning system, older adults’ attitude towards the new technology and most importantly the effect the system has on driving performance. Introducing a speech based in-vehicle system for older adults could potentially prove beneficial and improve driving performance and driver satisfaction, or it could prove disastrous and distract their attention from the driving task. The study presented in this chapter was designed to answer some of these questions.

4.3 In-Vehicle Hazard and Warning System for Older Adults

It is important to consider both the selection of voice and the format of the informational content when designing a speech-based in-vehicle system for older
Older Adult Drivers and Voices for In-Vehicle Systems

adults. Most older adults are found to be less able to absorb long instructions than younger people (Zajicek and Morrissey 2001). This supports the usefulness of an in-vehicle system with timely speech based traffic related information, since the information then appears at the point it is required and useful, instead of at the beginning of the driving task.

Humans are well-tuned to detect characteristics in a voice and use that skill when communicating with both humans and speech-based computer systems (Banse and Scherer 1996). Varying the linguistic and para-linguistic properties of the voices used by in-vehicle systems can also influence driving performance, as characteristics of voices can influence people’s attention and affect performance, judgment, and risk-taking (Jonsson, Nass et al. 2004; Nass and Brave 2005). Characteristics of the voice affects listeners perception of liking and credibility of what is said either by another human or by a speech-based system (Nass and Brave 2005). The psychological literature suggests that consistency is important. People expect consistency and prefer it to inconsistency. When inconsistency is encountered, people enter a state where they are motivated to adapt their perceptions in order to resolve inconsistency (Fiske and Taylor 1991). The need for consistency is well understood in traditional media, but is less clear for human-computer interaction. In the context of in-vehicle information systems, Nass et al.(Nass, Jonsson et al. 2005) clearly show a positive effect of matching the emotional characteristics of the in-vehicle voice with the emotional state of the driver, see Chapter 2. It has also been shown (Lazarsfeld and Merton 1948) that most successful human communication will occur between a source and a receiver who are alike, i.e., homophilous, and have a common frame of reference. Communication is more effective (Rogers and Bhowmik 1970) when source and receiver share common meanings, belief, and mutual understanding. Individuals enjoy the comfort of interacting with others who are similar. Matching personality (Nass and Lee 2000) and matching accents (Dahlbäck, Swamy et al. 2001) matter, people prefer system with personalities and accents similar to themselves. Talking with those who are markedly different from a person requires more effort to make communication effective, once again with a potential negative impact on driving performance. Theories of similarity-attraction and consistency-attraction (Byrne 1971) would also suggest that older adult driver would prefer an older adult voice for the in-vehicle system and that young people would prefer a young adult voice.

A study was designed to address the question of how voice selection and information content of an in-vehicle information system affects driving performance and attitude for drivers 55 years of age and older (Jonsson and Zajicek 2005). To minimize the cognitive load induced by the system, the in-vehicle system was designed to be a pure
information system providing hazard and warning messages. No interaction with the system was required, the system simply provided information on upcoming road hazards and traffic events without engaging the driver in a dialogue. Information by the system was designed to focus attention to traffic situations and by providing older adults with relevant information, extra time and distance to evaluate the situation, the hope was to improve their ability to react confidently. To investigate effects of similarity and consistency, the in-vehicle hazard and warning system was furthermore realized as two versions with the same information content but using different voices. One version had all warning and hazard messages presented by a young voice and the other version had all the messages presented by an older voice.

The in-vehicle hazard and warning system was storyboarded with a driving scenario and consisted of 25 speech prompts. The driving scenario was 60,000 feet (18.3 kilometer) and contained several hazards within a varied and realistic road scenario. The scenario took the drivers along a road that varied between one and two lanes in the drivers’ direction. There were small to medium sized villages placed along the road.

Note that the driving scenario was designed for left-hand side driving. The road varied between one and two lanes in the drivers’ direction.

The driving scenario took drivers past intersections, parked cars and pedestrians. The scenario was instrumented with moderate light conditions and moderate traffic.

Figure 4-1: Left hand side driving - two lane road
Most villages were small with only stop signs for intersections; larger villages had intersections with traffic lights, sometimes with multiple intersections and traffic lights. The density of pedestrians and parked cars varied between the villages. Based on the storyboard, obstacles such as road work, cars parked half-way off the road, and bicyclists, were placed along the road. Traffic was minimal to moderate. There were very few cars in the same direction as the participant driver and there were random and infrequent bursts of oncoming traffic.

The content of each of the in-vehicle system prompts was decided by the storyboard, based on obstacles and traffic events. Some hazard and warning prompts were of an ephemeral type that could feasibly be supplied by police reports or weather reports, other prompts were of a more permanent character and provided location of school zones and speed limits.

Drivers’ drove through a section of the scenario with dense fog. Hidden in the fog was an intersection controlled by a traffic-light. There were also numerous sections with road-work spread out along the road.
Speed limits varied between 20 and 60 miles per hour (32 and 97 kilometres per hour). Villages and school zones were areas with lower speed limits. There were objects obstructing the road. Depicted in Figure 4-4 is a fallen tree that blocks half the road in the drivers’ direction.

Ten of the prompts were designed to inform drivers of traffic events and road conditions, another 10 prompts were designed to warn drivers of upcoming hazards. Typical warning prompts would indicate traffic events, such as road work or slippery stretches, and road designs such as intersections and school zones that made it prudent for the drivers to slow down or to drive extra carefully. Even though the location of the in-vehicle system prompts was decided by the storyboard, the exact timing was tuned by test drivers. During the test drives, participants indicated when information should be given earlier or later. After five test drivers, the locations of the prompts were stable. There were also five prompts that suggested divertive routes to avoid potential hazards. Four of the suggestion prompts indicated that the driver should turn left or to turn right to avoid slow moving traffic or accidents. All prompts that suggested turns were proposed at intersections with traffic lights where the driver had to stop. The intent was to measure if drivers listened to the information prompts and whether drivers would take advice from the in-vehicle information system.

Special care was taken to make speech messages as short and succinct as possible. This resulted in the messages listed in Appendix C, exemplified by the following:

Warning prompts:

- Caution, there is thick fog ahead
- Warning there is a fallen tree in the road ahead.
- Beware of cyclists ahead
Information Prompts:

- You are approaching an intersection
- The current speed limit is 60 miles an hour
- Stop sign ahead

Suggestion Prompts

- The police use radar here, you might need to slow down
- There is heavy traffic ahead, if you turn left you can avoid it
- There is an accident ahead, if you turn right you might avoid it

Two voice talents, a 20 year old female and a 76 year old female, recorded all messages, using the exact scripted words in the message and speaking in a neutral and calm voice. Messages were reviewed by me and the voice talents during the recording session, and prompts were re-recorded until the two sets of prompts were matching in perceived amplitude, speech rate and tone of voice. To ensure that the voices were equal or comparable in quality and tone, and representative for a young voice and an older adult voice, the voices were then assessed by participants from two age groups, 18 to 25 and 55+. The assessment of the two voices of the in-vehicle hazard and warning system is described in Section 4.4. Section 4.5 describes the driving simulator study and in Section 4.6 there is a discussion on the results from the studies and how it affects design of in-vehicle information systems.

4.4 Assessing Two Voices for In-Vehicle System

The two voices used by the in-vehicle hazard and warning system were assessed in a separate study. In addition to ensure that the voices could be identified as a young voice and an older adult voice, participants were asked to assess the voices along measures as speech quality, intelligibility, emotional colouring, and homophily. This was done to control for differences that might otherwise bias the results when later used in the driving simulator study. More specifically we wanted to know the following:

- Are there noticeable differences in the characteristics of the two voices such as age of voice, emotional colouring, and quality of voice?
- Are there any differences in attitude towards the voices such as perceived trust and credibility?
- Are there any differences in how the two age groups identify themselves with two voices/persons speaking?
4.4.1 Design of Study

The experiment involved two gender balanced groups, one group with participants between the ages of 18 and 25, the other group with participants 55 years of age and older. Each participant assessed both sets of speech prompts for the in-vehicle information system, the young voice and the older adult voice. The order of assessing the voices was balanced between the participants in each group. Participants were informed that the study would last approximately 45 minutes and that they were expected to assess two voices by listening to the voices and filling in paper-and-pencil questionnaires. The experiment was conducted at Oxford Brookes University in the UK and all participants gave informed consent and were debriefed after the experiment.

Participants and Experimental Apparatus: There were a total of twelve participants, six male and six female, recruited for the study. Six participants were between 18 and 25 years of age (M= 22, SD =2.3 gender balanced), and six participants were 55 years of age and older (M=60, SD=4.4, gender balanced differences). The older group was recruited from the North Oxford area. Most of these subjects belonged to the local Philosophical Society and represented the group of older adults most likely to be interested in buying an in-vehicle messaging system. They all reported that they were in good health and currently held a driving license. The six younger subjects were students at Oxford Brookes University who were studying a wide variety of different subjects with a variety of levels of information technology competence.

The participants assessed the voices using a laptop with Microsoft Media Player and earphones. Participants controlled the laptop and could select and play any of the recorded prompts by clicking on them. The recordings consisted of the 25 short voice prompts for the in-vehicle information system, recorded in two different voices, a young female adult voice and an older female adult voice. For a list of prompts, please see Appendix C.

Procedure: Participants, young and old, were screened for their own perception of their hearing. It was explained that the experiments relied upon participants listening to speech messages and they were asked if they thought their hearing was adequate for this scenario. Two potential older participants withdrew as a result of doubts they expressed about their hearing.

Participants within each age group were randomly divided into two groups, one group that listened to and assessed the young voice first, and one group that listened to and
assessed the older adult voice first. All participants listened to all messages for the in-vehicle hazard and warning system, both the messages recorded by the young adult female (20 years old) speech talent and the messages recorded by the older adult female (76 years old) speech talent.

Participants were informed that they could play through the recorded speech prompts at any time and as often they liked while they filled in a set of paper-and-pencil questionnaires with their assessment of the voices. There were four questionnaires: one for emotional colouring of voice, one for trust and credibility of voice, one for quality of voice, and finally one for homophily (Rubin, Palmgreen et al. 1994), e.g. how the participants perceived the persons speaking being similar to themselves in terms of background and attitude as described below. When participants had finished assessing the first voice, they notified the experimenter and were provided with recordings and questionnaires to assess the second voice.

4.4.2 Measures for Assessing the In-Vehicle Voices
The data from the four questionnaires was used to create six measures. Three measures focused on perceived paralinguistic features and characteristics of voice, and three focused on perception of the voice or the speaker. All measures, except age of voice, consisted of indices created from sets of questions. The indices constructed for this study were tested for reliability using Cronbach’s alpha.

**Age of Voice:** An important question for the planned driving simulator study was the effect of age of voice. Participants were therefore asked to simply judge the age of the person speaking for both the young voice and for the older adult voice.

**Positive Emotional Colouring of Voice:** Positive Emotional Colouring of Voice of each voice was measured using a variation of the Differential Emotion Scale (DES) (Izard 1977). The DES scale was an abbreviated 20-term version divided over 5 subscales, resulting in 4 terms per emotion. The scale required participants to rate to which extent various emotional words were able to describe how they perceived the voice they listened using a 10-point Likert scales (1 = Describes very Poorly to 10 = Describes very Well). A positive emotion index was comprised of the eight terms happy, delighted, enthusiastic, amused, curious, attentive, alert and interested. The index was very reliable (Cronbach’s α = .77).

**Quality of Voice:** Quality of voice was measured using a questionnaire for which participants were asked to rate adjectives based on the question, “How well does each of the following words describe the voice you just heard?” The questionnaire included pairs of adjectives such as intelligible/inarticulate, loud/quiet, high pitch/low pitch,
fast/slow, breathy/non-breathy, hoarse/clear, hesitant/fluent, calm/stressed, and slurred/enunciated. Contrasting adjectives were paired on opposite sides of a 10-point Likert scale from 1 - 10. See Appendix C for a full list of contrasting pairs. The questionnaire was used to create one index, clarity of voice comprised of the terms intelligible, clear, non-breathy, fluent and enunciated. The index was very reliable (Cronbach’s α = .79).

Trust in Voice: Trust in voice was measured using a standard Individualized Trust questionnaire (Rubin, Palmgreen et al. 1994). Participants were asked to rate a number of adjectives based on the question “How well does each of the following words describe the voice you just heard?” Contrasting adjectives were paired on opposite sides of 10-point Likert scales from 1 - 10.

Credibility of Voice: Participants were asked to rate a set of adjectives based on the question, “How well does each of the following words describe the voice you just heard?” Contrasting adjectives were paired on opposite sides of 10-point Likert scales from 1 - 10. Standard measure of source credibility from Berlo’s and McCroskey’s credibility scales (Rubin, Palmgreen et al. 1994) were created. These measures, Authority, Character, Qualification, Dynamism and Safety, were combined into one index of source credibility.

Homophily: A standard questionnaire on homophily (Rubin, Palmgreen et al. 1994) was used to identify measures of similarity based on voice alone. Three indices were created, attitudinal similarity, behavioural similarity, and overall similarity. Participants were asked to rate the statements based on the question, “On the scales below, please indicate your feelings about the person speaking.” Contrasting statements were paired on opposite sides of 10-point scale from 1 – 10. Likert scales such that, for example, “similar to me” and “different from me” would appear at different ends.

4.4.3 Results
The assessments of the two voices focused on the perceived quality of the voices and a perception of the speakers based on their voices. The measures were analyzed using a two-way ANOVA’s with age group of participants (18-25, 55+) and gender as independent variables. Differences between the voices (young voice, older adult voice) were assessed using paired-samples T-tests.

Reported in the results below are effects to a significance level of .05. Trends – that is effects that are of interest but fails to reach significance - are also reported. Please note that M is short for Mean and SD is short for Standard Deviation.
Age of Voice: The young voice was perceived to be 18 years old by the young group, and 23 years old by the older adult group, \((F(1,8) = 5.3, p < .06)\), see Table 4-1 for mean values and standard deviation. Listening to the older adult voice, the young and older adult group both perceived the age of the speaker to be 63 years of age \((F(1,8) = .02, p < .91)\), see Table 4-1 for means and standard deviation. This placed the young voice, assessed as 18 and 23 year old (in reality 20 years old) within the age group 18-25. It also placed the older adult voice, assessed as 63 (in reality 76 years old) in the age range of 55+.

<table>
<thead>
<tr>
<th>Age of Voice</th>
<th>Young Voice</th>
<th>Older Voice</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age group</td>
<td>Mean</td>
<td>SD</td>
</tr>
<tr>
<td>18-25</td>
<td>18.33</td>
<td>2.7</td>
</tr>
<tr>
<td>55-73</td>
<td>23.2</td>
<td>4.8</td>
</tr>
<tr>
<td>Mean</td>
<td>22.96</td>
<td>4.23</td>
</tr>
<tr>
<td>SD</td>
<td>2.7</td>
<td>4.8</td>
</tr>
<tr>
<td>18-25</td>
<td>62.7</td>
<td>4.4</td>
</tr>
<tr>
<td>Mean</td>
<td>62.7</td>
<td>4.4</td>
</tr>
<tr>
<td>55-73</td>
<td>63.0</td>
<td>4.6</td>
</tr>
<tr>
<td>Mean</td>
<td>63.0</td>
<td>4.6</td>
</tr>
</tbody>
</table>

Table 4-1: Rating of Voices by Age Group

Emotional Colouring of Voice: The older adults and the young adults rated the voices quite differently. The older adults perceived both voices as significantly more positive (positive emotional colouring) than the younger group (older voice: \((F(1,8) = 12.7, p < .007)\); younger voice: \((F(1,8) = 26.4, p < .001)\), see Table 4-1. When comparing the two voices, both voices were considered similar in terms of emotional colouring, there was no significant difference \(t(11) = -2.2, p < .06\), see Table 4-2.

Quality of Voice: There were no significant differences between the perceived amplitude, pitch level, speech rate, and stress level of the two voices, nor was there any significant difference in how the two age groups rated the clarity of the young voice \((F(1,8) = .55, p < .48)\) and the clarity of the old voice \((F(1,8) = .08, p < .78)\). However, there was an evident main effect for voice, with both groups agreeing that
the older adult voice was more clear than the young voice, $t(11) = 3.0, p < .012$, see Table 4-2.

**Table 4-2: Comparison of Young Voice and Older Adult Voice.**

<table>
<thead>
<tr>
<th></th>
<th>Young Voice</th>
<th>Older Voice</th>
</tr>
</thead>
<tbody>
<tr>
<td>Positive emotional colouring</td>
<td>44.9</td>
<td>38.8</td>
</tr>
<tr>
<td>Clarity of voice (quality)</td>
<td>36.4</td>
<td>42.3</td>
</tr>
<tr>
<td>Trust of voice</td>
<td>28.5</td>
<td>47.3</td>
</tr>
<tr>
<td>Source Credibility</td>
<td>105.7</td>
<td>111.4</td>
</tr>
</tbody>
</table>

**Trust of Voice:** There was no significant difference in how the two age groups rated the trustworthiness of the two voices (older voice: $F(1,8) = .46, p<.52$; younger voice: $F(1,8) = .15 p>.71$), see Table 4-1. Both groups agreed that the older adult voice was significantly more trustworthy than the young voice $t(11) = 15.2, p < .001$, see Table 4-2.

**Credibility of Voice:** There were significant differences in the perceived credibility of the two voices. The older adults rated both the young voice ($F(1,8) = 13.7, p < .006$) and the older adult voice ($F(1,8) = 34.09, p < .001$) to be more credible than the younger adults, see Table 4-1. The older adult voice was also considered significantly more credible than the young voice $t(11) = 3.45, p < .005$, by both age groups, see Table 4-2.

**Homophily – Similarity:** The similarity measures showed significant differences in how the age groups rated the older adult voice. The older adult group felt more similar to the older adult voice (or rather person speaking) in both attitude ($F(1,8) = 45.2, p < .001$), and behaviour ($F(1,8) = 83.4, p < .001$) than the younger adult group, see Table 4-1. There were fewer significant differences when the two age groups rated the young voice. The young group perceived the young voice to be more similar to them in attitude ($F(1,8) = 6.4, p < .04$) than did the older group; however, there were no perceived differences in behavioural similarity ($F(1,8) = .0, p < .98$), see Table 4-1.

**4.4.4 Discussion**

The results from assessing and comparing the two voices showed that people could identify and reacted to cues from the recorded speech. The results on similarity show that the older adult group strongly identified themselves with the older adult voice. Young adults, however, only indicated similarity with the young voice on one of two measures of similarity. The assessment of age of voice is confirmed by the similarity scores, and participants in both age groups placed the age of the two voice talents
within the anticipated range of 18 to 25 and 55 to 73 respectively. Apart from this, the perceived differences between the two voices were that 1) the older adult voice was considered to be clearer than the younger voice, and 2) that the older adult voice was perceived to be more trustworthy and credible than the younger voice. Previous studies on credibility show that characteristics like age, sex, socioeconomic status, or expertness may affect the perceptions of credibility (Hovland, Janis et al. 1953; Bettinghaus 1968). Age is often associated with experience and wisdom and this can provide an explanation for higher credibility and trust scores for the older voice.

The results above show that the target groups could identify the age of the two voices - with a slight quality bias towards the older voice. These two voices were used as an older adult voice and a young voice in the in-vehicle hazard and warning system. The hypothesis formed by these results was that similarity-attraction would hold for older adult group, so that the older adults would prefer the older adult voice for the in-vehicle system. For the younger group, the assessment did not give any clear indication of preference, the trust, credibility, and quality assessment show a preference for the older voice, and one of two similarity measures show a preference for the younger voice. These results predict very different outcomes and it therefore becomes important to investigate which voice is preferred when driving.

4.5 Assessing the In-Vehicle Hazard and Warning System

To investigate if and how the age of voice when used in an in-vehicle hazard and warning system affects drivers attitude and driving performance, the two voices of the in-vehicle hazard and warning system (assessed in section 4.4) were tested in a driving simulator. The driving simulator was setup with three driving conditions all using the same storyboarded and scripted driving scenario; see section 4.3 for a description of the driving scenario and the speech prompts used for the in-vehicle hazard and warning system. One driving condition was setup with the older adult voice for the hazard and warning system, the second driving condition was setup with the young adult voice for the hazard and warning system, and the third driving conditions was setup without the hazard and warning system. These three driving condition where specifically designed to answer the following questions:

- Will an in-vehicle information and hazard warning system positively influence the driving performance for older adult drivers?
- Does the voice matter?
  - Will the age of the voice influence driving performance?
  - Will the age of the voice influence attitude and perception of the in-vehicle information system?
4.5.1 Design of Study

The experiment was a 2 (age group: 18 to 25 and 55+) by 3 (in-vehicle system: older adult voice, young voice, and no voice system) gender-balanced, between-participants design. The experiment consisted of a two week study at Oxford Brookes University, UK, using a driving simulator. Participants were informed that the experiment would last approximately one hour. All participants volunteered their time for the experiments, gave informed consent, and were debriefed at the end of the experiment.

The Driving Simulator with In-vehicle Information System: The driving simulator used in the experiment was STISIM Drive model 100 with a 45 degree driver field-of-view from Systems Technology Inc. Participants sat in a real car seat and “drove” using a Microsoft Sidewinder steering wheel and pedals consisting of accelerator and brake. The simulated journey was projected on a wall in front of participants and was set to daylight and cloudy but no rain, thereby avoiding bright light which might be distracting. All drivers used the same car, thereby experiencing the same vehicle properties such as acceleration, brakes, and traction. Furthermore, all drivers completed the same driving scenario (see section 4.3 for a description) even though they were assigned different conditions based on the in-vehicle information system. A driving scenario in STISIM Drive is static and predetermined; the scenario has a specific length and takes all drivers along the exact same road regardless of left and right turns. Based on this feature of STISIM Drive, all participants drove the same route for the same distance (60,000 feet, 18.2 km). The in-vehicle system, as described in Section 2 and assessed in Section 3 of this chapter, was played out of speakers in front of the driver, mimicking the sounds that would come from speakers on the dashboard. The amplitude of the in-vehicle system was set by two older adult test drivers because older people find it more difficult to distinguish speech in noisy environments (Gordon-Salent and Fitzgibbon 1999). This resulted in a setting that was noticeably louder than in previous driving experiments for age groups 18 to 25. The amplitude was kept at the same level for all participants.

Participants and Procedure: There were a total of 36 participants for the driving simulator study, 18 young adults (M = 22.1, SD 2.4) and 18 older adults (M = 63.1, SD = 5.9). Participants in this experiment were different people from those that carried out the voice only tests and were selected to match the two age groups (18 - 25, 55 and older) who participated in the voice only experiment. The eighteen younger subjects were students at Oxford Brookes University who were studying a wide variety of different subjects with a variety of levels of IT competence. The older group was recruited from the North Oxford area. Most of the subjects either belonged to local Philosophical Society or the local bell-ringers club. These subjects were
chosen as they represented the group of older adults most likely to be interested in buying an in-vehicle messaging system. They were relatively wealthy being from the high socio-economic level and were currently holding a driving license. All participants reported that they were in good health and (as demonstrated below) had similar driving habits.

As in the experiment assessing the two voices potential participants, young and old, were screened based on their perception of their hearing. It was explained that the experiments relied upon participants listening to speech messages in a car environment implemented by a driving simulator and they were asked if they thought their hearing was adequate for this scenario. Three potential older participants withdrew as a result of doubts they expressed about their hearing.

Participants in the older adult group were randomly placed in three gender-balanced and age-balanced groups:

- **Group 1** – driving setup 1 with in-vehicle hazard and warning system using young voice.
- **Group 2** – driving setup 2 with in-vehicle hazard and warning system using older adult voice.
- **Group 3** – driving setup 3 without an in-vehicle system.

The same procedure was used to place the young adults in the three groups:

- **Group 4** – driving setup 1 with in-vehicle hazard and warning system using young voice.
- **Group 5** – driving setup 2 with in-vehicle hazard and warning system using older adult voice.
- **Group 6** – driving setup 3 without an in-vehicle system.

Participants completed a set of pre-driving questionnaires. Older adults filled in three questionnaires: 1) self-report on driving habits, accidents and tickets during the past years, 2) self-report on perceived driving abilities based on a questionnaire for drivers 55 and older from the AAA Foundation for Traffic Safety (AAA 1994); and 3) a fifteen-term Differential Emotional Scale (DES) (Izard 1977) used to record the participants’ current emotional status. Young participants filled in questionnaires 1 and 3.

Each participant drove a test run to familiarize themselves with the workings of the simulator. This is particularly important for older adult drivers since previous studies
show that older adults need about three minutes of driving to adapt and get the feel of the driving simulator (McGehee, Lee et al. 2001). All participants experienced a car crash in the simulator because this can otherwise be a stressful and uncomfortable experience when the simulator generates a loud crash, the windshield breaks, and the simulation halts for a couple of seconds. After the crash, most participants let go of the steering wheel and leaned back in the car seat after an accident, and when the simulator started again they were ready to continue driving. For this experiment we used a five minute training course to enable participants to have a normal driving experience and to screen for participants with simulator sickness (Bertin, Guillot et al. 2004). There were three participants from the older adult group that did not finish the training course and had to withdraw from the experiment due to nausea, three additional older adult participants were recruited.

After the training course participants were instructed that the driving session would take approximately 30 minutes and that they should drive normally and obey normal traffic rules and posted speeds limits. Groups driving with the in-vehicle hazard and warning system were told that there would be a speech based system in the car to inform them about traffic and road conditions in a recorded female voice.

After driving, each participant completed two questionnaires. The first post-task questionnaire was the same fifteen-term DES self-report as used before driving, and the second questionnaire asked participants how they felt about the credibility and quality of the in-vehicle hazard and warning system.

4.5.2 Measures for Driving Simulator Experiment

Data was collected for three pre-driving measures, age, driving habits, and driving performance. In addition to this, there were three measures for participants’ attitude and perception of the in-vehicle hazard and warning system, and three measures for driving performance. Some measures were based on standard questionnaires and groupings, and some measures were suggested by factor analysis; the reliability of these indices was determined by Cronbach’s $\alpha$.

**Age of participant:** Self-reported via questionnaire

**Driving Habits and Real World Driving Performance:** Participants self-reported on normal driving behaviour and incidents during the past two years. Measures were created based on the following questions: miles per week driven, driving with passengers, how well participants liked driving, number of accidents and number of tickets.
Positive Emotional State: The positive emotional state of each participant was measured before and after the driving session. Two positive emotion indices were created based on the DES questionnaire (Izard, 1977) and 10-point Likert scales (1=Describes very Poorly to 10 = Describes very Well). The index was comprised of the terms calm, relaxed, at ease, excited and happy. The indices were reliable (Cronbach’s α = .70 and .80, respectively). The difference between the before- and after-measures of positive emotional state was used to detect changes in emotions after driving.

Source Credibility of In-vehicle Hazard and Warning System: The credibility of the voice system was assessed by two standards scales for source credibility, McCroskey’s and Berlo’s (Rubin, Palmgreen et al. 1994). Participants were asked to rate adjectives based on their views of the system. Contrasting adjectives were paired on opposite sides of a 10-point Likert scale from 1 - 10. We combined the standard indices for authoritativeness, character, qualification, safety, and dynamism into one index of source credibility. The index was reliable (Cronbach’s α = .75).

Quality of the In-vehicle Hazard and Warning System: Quality and value of the in-vehicle system was based on an index created based on responses to the question, “How well do each of these adjectives describe the in-vehicle system?” Adjectives were presented with 10-point Likert scales anchored by “Describes Very Poorly” (=1) and “Describes Very Well” (=10). The index was comprised of the items, “Fun to use”, “High in quality”, “I would use again”, “I would discourage family and friends from using it” (reverse coded) and “I would not buy it or pay for it” (reverse coded). The index was very reliable (Cronbach’s α = .84).

Driving Performance: Driving performance during the experiment was created from data generated by the driving simulator. The driving behaviours recorded were: 1) different types of accidents, such as off-road accidents when the driver hits an obstacle off the road, collisions with vehicles and obstacles on the road such as road work bollards; 2) swerving, where the driver crosses the lane boundaries, including centreline and road edge; 3) speed measured in miles per hour; and 4) time to finish the driving course (measured in seconds). Three measures were created from this data. The index of Accidents was a combination of collisions, off-road accidents and pedestrian accidents. The index for Swerving combined centreline crossings and road-edge excursions. The index for Driving Time measured the time in seconds to finish the driving course. Driving speed was used to check if drivers were exceeding the speed limit.
Influence of the voice system on driving performance: This measure, based on questionnaires for drivers 55 years of age and older (AAA, 1994), was generated by the question “The Voice made me:” followed by a list of statements based on 10-point Likert scale (1 = Describes very Poorly to 10 = Describes very Well). The index was comprised of “Watch more carefully at intersections”, “React faster to dangerous driving situations”, “More comfortable driving at faster speeds” and “A better driver in low visibility conditions”. The index was very reliable (Cronbach’s $\alpha = .93$)

4.5.3 Results

The effects of the voices of the in-vehicle information system, when used in a driving simulator, on attitude and driving performance were evaluated by a two-way ANOVA with the age of voice of the in-vehicle hazard and warning system and age group of drivers as between-participant factors. Comparison of characteristics of the two voices was evaluated by paired-sample T-tests.

Driving Habits and Status before Driving: Participants were well age-balanced across conditions for both the older adult age group and for the younger age group. There were no significant differences reported in driving habits and driving performance between any of the six conditions.

Table 4-3: Comparing Older Adult Drivers Prior to Driving with In-vehicle System

<table>
<thead>
<tr>
<th>Age of Participants</th>
<th>Young Voice</th>
<th>Older Voice</th>
<th>No Voice</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
</tr>
<tr>
<td>Tickets</td>
<td>66.3</td>
<td>6.31</td>
<td>62.7</td>
</tr>
<tr>
<td>Accidents</td>
<td>.33</td>
<td>.516</td>
<td>.17</td>
</tr>
</tbody>
</table>

There were no significant difference in the number of tickets received across conditions, $F(2, 30) = .64, p < .53$, or between age groups, $F(1, 30) = 1.4, p < .24$. The data showed no significant difference in the number of accidents across conditions $F(2, 30) = .18, p < .84)$. Young drivers did however report involvement in significantly more accidents than older adult drivers, $F(2,30) = 6.4, p < .02$. This might also be due to the fact that only accidents during the past two years were reported, not lifetime accidents. Even though the difference is significant, there were very few reported accidents overall, with 94% of the older adults reporting no accidents and 61% of the young people reporting no accidents, see Table 4-3 and Table 4-4.
Table 4-4: Comparing Young Adult Groups Prior to Driving with In-vehicle System

<table>
<thead>
<tr>
<th>Age of Participants</th>
<th>Young Voice</th>
<th>Older Voice</th>
<th>No Voice</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
<td>SD</td>
</tr>
<tr>
<td>Age of Participants</td>
<td>22.8</td>
<td>2.2</td>
<td>22.8</td>
</tr>
<tr>
<td>Tickets</td>
<td>.17</td>
<td>.408</td>
<td>.00</td>
</tr>
<tr>
<td>Accidents</td>
<td>.50</td>
<td>.548</td>
<td>.17</td>
</tr>
</tbody>
</table>

Participants from both age groups reported no significant differences in driving habits such as area for driving (city, rural etc.), $F(2,30) = 1.9, p < .16$ or distance driven per week, $F(2, 30) = 1.0, p < .38$.

Positive Emotional State Before and After Driving: The data from the self-report of positive emotional state showed no differences across conditions before driving, $F(2,30) = 1.3, p > .29$. There was no difference in emotional state between age groups before driving, $F(1, 30) = .002, p > .97$, see Table 4-5.

After driving, all older adults experienced feelings of being less calm and at ease than before driving. They felt significantly worse than younger drivers $F(1,30) = 7.0, p < .01$, see Table 4-5. There was also an effect of voice system, where all drivers with the young voice felt more relaxed than drivers in the other conditions, $F(2,30) = 3.9, p < .03$. Drivers with the old voice felt worse, drivers with the young voice felt best, with drivers without a system in the middle.

Table 4-5: Comparing All Six Groups, Emotional State before and After Driving

<table>
<thead>
<tr>
<th>Positive Emotional State before driving</th>
<th>Young Voice</th>
<th>Older Voice</th>
<th>No Voice</th>
</tr>
</thead>
<tbody>
<tr>
<td>Olders Drivers</td>
<td>37.67</td>
<td>7.79</td>
<td>28.5</td>
</tr>
<tr>
<td>Young Drivers</td>
<td>33.67</td>
<td>4.55</td>
<td>32.50</td>
</tr>
<tr>
<td>Difference in positive emotional state before and after driving</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Older Drivers</td>
<td>-5.27</td>
<td>6.64</td>
<td>-18.95</td>
</tr>
<tr>
<td>Young Drivers</td>
<td>0.5</td>
<td>0.77</td>
<td>-6.20</td>
</tr>
</tbody>
</table>
The young voice had a good influence on older adult drivers. Analyzing the data from the older adult drivers using a one-way ANOVA, the results show that drivers with young voice system were able to maintain a more positive emotional state ($F(2,15) = 4.4, p < .03$, see Table 4-5) than drivers in the other two conditions. Drivers that interacted with the in-vehicle system with the older adult voice, showed the largest significant decrease in positive emotions of all conditions ($F(2, 30) = 9.2, p < .001$, see Table 4-5). They felt worse than drivers with the young voice and drivers without a system.

Analyzing the young drivers data with a one-way ANOVA, the data show that the voice had very little effect at all on emotional state before and after driving ($F(2,15) = .49, p < .62$, see Table 4-5). There was however a trend that showed a decrease in positive emotional state after driving with the older adult voice system ($F(2,15) = 3.3, p < .06$), see Table 4-5. This needs to be further investigated.

**Source Credibility of In-vehicle Information System:** Young adult and older adult drivers rated the source credibility of the two voices quite differently. There was no main effect for voice system, $F(1,20) = .96, p < .34$, or for age group, $F(1,20) = 1.8, p < .19$.

<table>
<thead>
<tr>
<th>Source Credibility</th>
<th>Young Voice</th>
<th>Older Voice</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
</tr>
<tr>
<td>Older Drivers</td>
<td>155.47</td>
<td>5.10</td>
</tr>
<tr>
<td>Young Drivers</td>
<td>132.90</td>
<td>11.7</td>
</tr>
<tr>
<td>Quality of In-vehicle System</td>
<td>35.73</td>
<td>11.9</td>
</tr>
<tr>
<td>Older Drivers</td>
<td>35.67</td>
<td>8.66</td>
</tr>
<tr>
<td>Young Drivers</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

There was however an interaction effect, $F(1,20) = 8.0, p < .01$. The data show that even though both young adults and older adults assessed the source credibility of the older adult voice to be the same, they did not agree on the assessment of the young voice. Older drivers rated the source credibility of the young voice higher than the source credibility of the older adult voice. Young drivers indicated no significant difference in source credibility between the young and the older adult voice used by the in-vehicle system.

Please note that this result, that the younger voice is seen as equal to (or even more credible than) the older adult voice, contradicts the assessment of the two voices in
the office setting, in which both age groups agreed that the older adult voice was more credible than the young voice.

**Quality of In-vehicle Information System:** There was no perceived significant difference in quality between the two systems, $F(1,20) = 1.8, p < .19$. There was however a difference in how the age groups assessed the quality of the voice system, $F(1,20) = 5.8, p < .02$. Older drivers rated the quality of the voice systems higher than younger drivers.

There was also an interaction effect. Even though the information presented by both voices was identical, older adult drivers perceived that the in-vehicle system with the young voice was more valuable and of better quality than the in-vehicle system with the older adult voice, $F(1,20) = 5.9, p < 0.02$, see Table 4-6.

Young drivers did not perceive there to be any significant difference in the quality of the in-vehicle system based on the different voices.

**Driving Performance Older Adult Drivers:** The speech messages of the in-vehicle hazard and warning system were placed right before an event or a hazard; the placement and the content of warning prompts caused most drivers to slow down.

More than 50% of the accidents experienced by older adult drivers with the in-vehicle system happened when avoiding a tree that had fallen across the road. These were low speed accidents mostly due to difficulty manoeuvring the car around the obstacle. Eighty percent of the remaining accidents were not associated with any speech prompt.

Older adults driving without the in-vehicle system had 68% of the accidents at locations where the in-vehicle system would have warned the driver.

Older adults showed significantly better driving performance on two out of three driving performance measures when driving with the in-vehicle hazard and warning system than when driving without the in-vehicle system.

There was a significant reduction in accidents ($F(1,32) = 15.1, p < .001$) and in swerving ($F(1,32) = 19.8, p < .001$), see Table 4-7. There was no difference in time to finish the course ($F(1,16) = 0.38, p > .54$, see Table 4-7).
Table 4-7: Driving Performance with or without In-vehicle System

<table>
<thead>
<tr>
<th></th>
<th>In-vehicle Information System (Young Voice and Older Voice)</th>
<th>No In-vehicle Information System (No Voice)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
</tr>
<tr>
<td><strong>Accidents</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Older Drivers</td>
<td>1.68</td>
<td>1.16</td>
</tr>
<tr>
<td>Young Drivers</td>
<td>2.25</td>
<td>1.60</td>
</tr>
<tr>
<td><strong>Swerving</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Older Drivers</td>
<td>20.45</td>
<td>4.34</td>
</tr>
<tr>
<td>Young Drivers</td>
<td>26.02</td>
<td>13.16</td>
</tr>
<tr>
<td><strong>Driving time</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(in seconds)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Older Drivers</td>
<td>1859</td>
<td>320</td>
</tr>
<tr>
<td>Young Drivers</td>
<td>1233</td>
<td>190</td>
</tr>
</tbody>
</table>

Investigating the effect of the two voices, the data show that older adults benefited from driving with a young adult voice giving them hazard and warning information. Older adults drove significantly better when driving with the young voice system than driving with the system with the older adult voice system. They had significantly fewer accidents, $F(2, 30) = 7.76, p < .002$, and they also completed the driving course faster than when driving with the older adult voice system or without a system, $F(2,30) = 13.1, p < .001$, see Table 4-8.

Table 4-8: Driving Performance with Young Voice and Older Adult Voice

<table>
<thead>
<tr>
<th></th>
<th>Young Voice</th>
<th>Older Voice</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
</tr>
<tr>
<td><strong>Accidents</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Older Drivers</td>
<td>1.17</td>
<td>.75</td>
</tr>
<tr>
<td>Young Drivers</td>
<td>2.17</td>
<td>1.47</td>
</tr>
<tr>
<td><strong>Swerving</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Older Drivers</td>
<td>22.7</td>
<td>2.39</td>
</tr>
<tr>
<td>Young Drivers</td>
<td>14.6</td>
<td>2.28</td>
</tr>
<tr>
<td><strong>Driving time</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(in seconds)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Older Drivers</td>
<td>1570</td>
<td>65.1</td>
</tr>
<tr>
<td>Young Drivers</td>
<td>1175</td>
<td>241</td>
</tr>
</tbody>
</table>

This is an interesting result since it indicates that even though they were driving faster, without exceeding the speed limit, they had fewer incidents than drivers in the other conditions (drivers with the older adult voice or drivers without a system).

**Driving Performance Young Drivers:** More than 95% of the younger drivers exceeded the speed limits multiple times. Most of the incidents experienced by the younger drivers were related to speed exceedance. As opposed to the older drivers, the young drivers were good at manoeuvring around the fallen tree. Similar to the
Older adult drivers, the younger drivers with in-vehicle system had 75% of their accidents in locations where no speech prompt had been issued, and young drivers without the system had 70% of their accidents where the in-vehicle hazard and warning system was designed to give a warning prompt.

It is interesting to note that young drivers benefited as much from driving with the in-vehicle system as the older adults. Young drivers also had significantly fewer accidents driving with the system than without the system, $F(1,32) = 15.0, p < .001$, see Table 4-7, and they swerved considerably less, $F(1, 32) = 19.8, p < .001$. There was no difference in time completing the driving course, $F(1,32) = .38, p < .54$ (see Table 4-7). The improvement in driving performance reported here, comes from a 100% accurate in-vehicle system, and confirms the improvement in driving performance for young drivers when driving with an in-vehicle hazard and warning system as presented in chapter 3.

The age of the voice of the in-vehicle information system seemed to have very little influence on incidents caused by young drivers, as opposed to the clear effect the age of voice had on older adult drivers. Swerving was the only driving performance factor that was influenced by the age of voice of the in-vehicle system: Young drivers swerved significantly less with the young voice in the in-vehicle system than with the older adult voice system, $F(2,30)= 29.0, p < .001$, see Table 4-8.

**Influence of System and Reactions to System**: Analyzing the data from the older drivers with a one-way ANOVA showed that drivers perceived that the system with the young voice had a positive influence on their driving performance compared with the system with the older adult voice, $F(1,10) = 31.38, p < 0.001$ (see Table 4-9). This index was only compiled for the older adult drivers since it was based on a questionnaire targeted for drivers 55 years of age and older.

<table>
<thead>
<tr>
<th>Influence of Voice System on driving performance</th>
<th>Young Voice</th>
<th>Older Voice</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
</tr>
<tr>
<td>Older Drivers</td>
<td>27.1</td>
<td>3.35</td>
</tr>
</tbody>
</table>

The in-vehicle information system presented four prompts suggesting that the driver should turn left or right. All of these prompts were located at intersections with traffic lights where the driver had to stop. Even if the drivers did not have more than a 45 degree view, this was offset by the fact that the intersections where they were asked to
turn left or right were controlled by traffic lights and that oncoming and crossing traffic was programmed to follow traffic rules, i.e., to stop at traffic lights when they show a red light.

The data from the driving sessions show that participants tended to follow suggestions from the in-vehicle system. The first recommended turn was missed by one participant, an older female driver driving with the older adult voice. The second recommended turn was followed by all participants. The third recommended turn was missed by two participants, both young male drivers driving with the young voice. Finally, the fourth turn was missed by three participants, the same two young male drivers that missed the third turn, and a third young male driver driving with the older adult voice. Only six young drivers elected to make turns without being prompted to do so.

4.5.4 Discussion

Assessment of the older adult voice and the young voice when people listen to the recorded voices using a laptop in an office setting shows that the older adult voice is more credible, more trustworthy, and more intelligible than the young voice. This assessment of trust and credibility changes dramatically when the older adult voice is used by an in-vehicle hazard and warning system in a driving simulator. The hypothesis based on similarity attraction and homophily - that the older adult voice would be better liked and trusted by older adults driving with the in-vehicle hazard and warning system - was contradicted by the results from the driving simulator experiment. The data from the driving simulator experiment instead show that the driving performance of older adult drivers was significantly better when driving with the young voice than the older adult voice. This result is very interesting and does not conform to most other studies on social responses where similarity attraction effects have been shown to hold.

First, however, we have to rule out bias in the realization of components of the study. One is that the young voice had better quality and was easier to understand. The data from the voice assessment, however, instead show that older adult voice was considered being of marginally better quality and more intelligible. Another possible explanation is that the younger voice is preferred a result of homophily, a perception of similarity. The data from the homophily questionnaire, however, clearly show that older adult drivers identified with the older adult voice, sharing common beliefs and a common frame of reference. We are left with two different environments, the office setting and the car, and clearly the car presents a different environment where other considerations should be taken into account. The car presents an environment where cognitive abilities (quick decisions, reaction times, taking in complex situations) and
physical abilities (hearing, vision and strength) are more in demand than the office setting. As a result of this, we hypothesize that the older adult drivers possibly projected their own declining physical and attention abilities to the voice of the in-vehicle information system. They might therefore trust an older adult voice less than a young voice in the car, as the young voice could be associated with properties of young adults such as better vision, attention, and cognitive abilities. If this is true, would this hold only for driving? Or is this a pattern of collaboration and communication that spans other domains where cognitive and physical abilities associated with younger age is required? This hypothesis requires more investigation, and should be extended to cover more domains that require quick real-time decisions.

Please note that regardless of the validity of the above posed hypothesis, the findings from this study highlight the importance of context and emphasize that a voice system in the car is different from a voice system on a computer in an office. Hence, in-vehicle systems or products targeted for the car should, for reliability of results, be tested in a driving simulator and ultimately in a real car (Jonsson and Zajicek 2005).

This experiment also demonstrates that there is significant potential for increasing the safety of older adult drivers (over 55 years of age) and also for young drivers (Jonsson, Nass et al. 2005) by providing timely information concerning road hazard. Results also show that the information is well-received by the drivers if the voice of the system is carefully selected. It is clear that the choice of voice for an in-vehicle information system is very important, and especially so for older adult drivers. The judgments of older participants in this study were far more affected by a change in context than those of younger people. Dulude (2002) found more flexibility in younger people when investigating performance with interactive voice response systems. These results point to the older population as a very different user group from younger people and also to the need for one to be aware of these age differences in every step of the evaluation process. The results also indicate that the car represents a truly different setting for information systems. Voices must be tested and selected in the relevant context: properties in a voice-only setting might be perceived differently in a driving simulator or a car than in other settings, resulting in unexpected influences on driving performance. It would therefore be advisable to test the voice intended for use in the car in a driving simulator or car, before proceeding to productize the voice for an in-vehicle information system.

The most striking effect on driving performance by the in-vehicle hazard and warning system was the data that showed that older adult drivers with the young voice finished the driving course much faster than older adult drivers in the other conditions (see Table 4-8). This increased average speed was done without exceeding the speed limit.
and with fewer incidents signifying that drivers in this condition felt more comfortable and confident driving at a higher speed. This is a positive effect for many reasons. It reduces the time to reach a destination, which has the potential to boost attitude (confidence and performance related). It reduces the risk of some accidents, since the lower speed of many older adult drivers tends to cause other drivers to take (unnecessary) risks in order to overtake them.

A limitation of the experiments described above is that they were carried out with a small number of subjects. One might expect greater diversity especially among older adults in the way that they experience age-related impairments. However the average within-group variation for older adults in Tables 4-1 – 4-9 shows the groups to be extremely homogeneous with a low standard deviation in their responses. As described in Sections 4.4.1 and 4.5.1, the participants were carefully filtered to represent only those in both age groups who would be able to respond both economically and educationally to the provision of the in-vehicle hazard and warning messages. Furthermore, Hakamies-Blomqvist and Henriksson (Hakamies-Blomqvist and Henriksson 1999) argue that age-related changes that affect driving usually do not become marked before the age of seventy-five. The older group in this experiment was in the range 55 to 73, and we would hence not expect there to be to great diversity within the group.

4.6 Conclusions

The goals with these experiments were to learn more about the effects of an in-vehicle information system on older adults, and to learn about the effect of the system’s voice. Data from the experiments give clear indications on how driver attitude and driver performance is affected by both voice characteristics and style of in-vehicle information system.

The data shows that an in-vehicle hazard and warning system can have a positive influence of the driving performance of older adult drivers. Results show that driving safety for older adults was enhanced with the use of an in-vehicle hazard and warning systems with a young adult voice (see Table 4-8). Older adults also perceived that the in-vehicle hazard and warning system with the younger voice had a positive influence on their driving (see Table 4-9). Investigating if the age of the voice selected for the in-vehicle hazard and warning system affected drivers, the data show that older adult drivers found the young voice to be more credible than the older adult voice (see Table 4-6). They also felt more at-ease after driving with the young voice than with the older adult voice (see Table 4-5). It is interesting to note that the age of the voice for an in-vehicle hazard and warning seemed to matter less for younger adults (see Tables 4-5, 4-6 and 4-8).
Older Adult Drivers and Voices for In-Vehicle Systems

The results on quality and influence of voice used by the in-vehicle hazard and warning system show that age of the voice can influence attitude and perception of the in-vehicle system. Older adults clearly associated high quality and perceived positive influence of the in-vehicle hazard and warning system with a young voice. Once again, the data show that younger drivers did not perceive that the age of the voice had any influence on the quality of the system.

This experiment demonstrates that there is significant potential for increasing the safety of both older adult drivers (over 55 years of age) and young drivers by providing timely information concerning road hazard. Results also show that the information is well-received by the drivers if the voice of the system is carefully selected. It is clear that the choice of voice for an in-vehicle information system is very important, and especially so for older adult drivers. Information provided by the in-vehicle information system should also be designed to assist older adult drivers with driving tasks associated with high risk factors. The information and hazard warning system presented in this work provided support for only two of the seven risk factors listed in section 1 of this chapter. First, failure to maintain proper speed was addressed since older adults often drive too slow and one of the outcomes of the experiment shows that older adults drove faster with the system than without the system (without exceeding the speed limit), and second, inattention was addressed by the system that directed the drivers’ attention to upcoming traffic events and road hazards. More of the risk factors can be addressed using additional information on road designs, road conditions, and hazards currently available from sources such as police, civil authorities and other cars. This information can when relayed to drivers, potentially make them even better prepared for the road ahead. For further work it would prove useful to investigate the effects of providing more categories of information in the in-vehicle system to address some of the remaining risk factors for older adult drivers as listed in Section 4.1.

It is important to investigate how drivers react to different types of information content, especially suggestions and proposals made by an in-vehicle information system. Five of the information prompts in the in-vehicle information system in this experiment were phrased as suggestions, such as “The police use radar here, you might need to slow down” and “There is a traffic congestion ahead, if you turn right you can avoid it”. These were inserted to measure how drivers listened to and trusted suggestions made by the system. The data show that all but four drivers (of 36), regardless of voice, drove according to the suggestions of the in-vehicle system. This led us to believe that drivers, regardless of age, listened to and heard the information prompts. However, further work is needed to investigate what type of suggestions and
proposals drivers follow, if there are different patterns for older and younger drivers, and if compliance is limited to suggestions about traffic.

The judgments of older participants in this study were far more affected by a change in context than those of younger people. Dulude (2002) show that performance was worse for older people than younger users, because older people responded more negatively to design problems that made their interaction difficult, whereas younger people were more flexible and able to work around them. Data from this study confirms this finding because the judgment of voices by the younger group was more stable between the voice only and the driving simulator and more independent of context than the judgment of voice by the older adult group. These results point to the need for a contextual focus in interface evaluation and a recognition of the value of the design in a wider context. Cockton (2005) argues that the focus of interface design has shifted over the years from the system via the user to the context of use, and that all are necessary but not sufficient for effective interactive systems design which requires a new value-cantered focus. In his view ‘the value-cantered framework involves opportunity identification as well as design, evaluation and iteration where opportunity identification has the goal of stating the intended value for a digital produce or service in the world’. The evaluation of the speech-based in-vehicle hazard and warning system presented here involved assessing value. It is however in the nature of safety critical systems, such as in-vehicle systems, that this value must be perceived or evaluated in simulation of the safety critical scenario.

The reported results from a study conducted in a driving simulator, indicates that further investigations should be done with real cars to verify possibilities for increasing safety on the roads. This is true for both older adult and young drivers, but as the results show, more important for older adult drivers. These results once again point to the older population as a very different user group from younger people and also to the need for one to be aware of these age differences in every step of the evaluation process. The results also indicate that the car represents a truly different setting for information systems. Voices must be tested and selected in the relevant context: properties in a voice-only setting might be perceived differently in a driving simulator or a car than in other settings, resulting in unexpected influences on driving performance. It would therefore be advisable to test the voice intended for use in the car in a driving simulator or car, before proceeding to productize the voice for an in-vehicle information system. Assessment of how technology affects older adult drivers also often lack real word data (Coughlin 2003). Even though surveys, sensor information from real car, and as in this study, simulated driving, provide powerful insights into how new in-vehicle systems affect older drivers, many crucial effects only emerge through prolonged used by a larger population.
4.7 Acknowledgements and Publications

I designed the study, the driving course and the initial prompts. I selected the questionnaires, some of the standard. I adapted the pre-and post-driving questionnaire from Stanford University to match the study. Mary Zajicek, Oxford Brookes University, adapted the hazard and warning prompts from US-English to UK-English, and also adapted them to suit older adults. Students from Oxford Brookes helped run the study, and I analyzed the data.

Material from this study has been published as follows:


Zajicek, M. and I. M. Jonsson (2005). In-car Speech Systems for Older Adults: Can they Help and Does the voice Matter Technology, Knowledge and Society, Hyderabad, India.


Jonsson, I.-M., M. Zajicek, et al. (2005). In-car Speech Based Information Systems for Older Adults. 12th World Congress on Intelligent Transport Systems and Services, San Francisco, CA, USA.


Older Adult Drivers and Voices for In-Vehicle Systems
5 Summary, Discussion and Conclusions
This chapter contains a summary of the results from chapters 2, 3 and 4 followed by short descriptions of additional studies of how various properties of voices used by in-vehicle systems impact drivers’ performance and attitude. Most importantly we describe how results from these studies confirm or add to the results from the studies described in previous chapters. Finally we summarize all studies with a discussion on limitations, generalizations and cultural considerations.

5.1 Summary of Results from Previous Chapters
The key finding from the studies presented in chapters 2 to 4 is that using a voice in the car is different from presenting text and triggers cultural and social effects. In addition to this, the studies presented here all show that characteristics of the voice used in the car can influence driving performance. A well selected voice can have a positive influence on driving performance whereas an inappropriate voice can negatively impact both driving performance and attitude. Most importantly, the studies clearly show that there is not one voice that is effective for all drivers.

As observed in chapter 1, a voice conveys more than just the content of what is spoken. As listeners we can identify a number of general speaker characteristics, such as age and sex, as well as a number of social factors including ethnic background, educational level, and social class. Even though there might be cultural and social differences in how large the linguistic differences are between different groups, there are two common themes. First, people identify speaker characteristics automatically and without effort, even to the point that it is something that they cannot avoid doing. Second, as listeners we are influenced in our interpretation of a message by our attitude towards the speaker. It would be reasonable to assume that voice communication with an in-vehicle system car would be influenced by these same factors.

The studies presented here clearly demonstrate that voices from in-vehicle systems are associated with social cues and triggers that influence perception of conveyed messages. Not only was the uptake of the presented information influenced by these factors, but the driving behaviour was affected by variations in the voices characteristics. More importantly, the data also demonstrate that drivers benefit from an in-vehicle information system that knows its driver. This suggests that in-vehicle systems should be carefully designed, and that voices used by the system should be carefully selected.
The attitudinal measures used in the studies focus on trust, liking and homophily, and predict (often based on similarity attraction) drivers comfort and willingness to communicate with the in-vehicle system. Since all studies presented in this thesis were conducted in a driving simulator, it enabled us to collect behavioural and performance measures in addition to attitudinal measures. The behavioural and performance measures were based on monitoring driver’s attention to the driving task, either via direct driving actions such as brake and accelerator usage, number of accidents and adherence to traffic regulation, or indirect responses to horn-honks. Most importantly, the added behaviour and performance measures enable us to investigate correlations between attitude, perceived performance and measured performance. It is interesting (but not surprising) to note that the attitudinal measures are linked to the behavioural measures. This confirms the literature that happy and content people tend to perform better than sad and discontent people (Groeger 2000).

Most studies presented here confirm attitudinal findings from social sciences. Familiarity (with positive association), and accuracy both work to promote trust and liking. The study on emotional drivers confirms previous findings in social studies that corroborate advantages of similarity attraction. People tend to like voices that have one or more properties and/or characteristics that match their own voice. Performance data show that drivers that liked and trusted the in-vehicle system always (on average) showed better driving performance than drivers that disliked and distrusted the in-vehicle system. A similar spill-over effect (halo effect) can also be found in drivers’ perception of the car. This spill-over effect predicts that if drivers like the in-vehicle system they also like the car, and vice versa. Furthermore, there is also a risk that a new in-vehicle system could affect any existing in-vehicle systems in addition to the car. For example; if a driver dislikes a new in-vehicle hazard and warning system, this could negatively influence the trust in other systems in the car such as navigation systems or car-status information systems. Dislike and distrust in one system have the potential to reduce the credibility of all systems as well as the car itself. This trend was noted in all studies described in chapters 2 to 4, and is a note to automobile manufacturers to think carefully about branding when introducing in-vehicle systems.

Even though most studies presented in chapters 2 to 4 confirm previous findings in the literature, there is one study in particular where the data contradict findings of similarity attraction. Data from the age-of-voice study presented in chapter 4 show that even though older adults like and trust an older voice in the office, they prefer a young voice in the car. This finding indicates that context matters, and that driving makes people react to voices differently than when interacting with a computer in the office. It is clear that this result does not conform to and confirm similarity attraction,
Social and Emotional Characteristics of Speech-based In-Vehicle Information Systems: Impact on Attitude and Driving Behaviour

an otherwise solid finding in social sciences. The data also show that this affects performance, so that drivers with a young voice system exhibit better driving performance than drivers with an older voice system. What makes the driving environment different? Is this reaction confined to age-of-voice and older adult drivers? Or are there other properties and characteristics of voice that produce the same unexpected results?

In addition the work described in detail in this thesis, I have in collaboration with co-workers conducted more studies in this field. These studies focused mainly on trait properties of voice such as age and personality, and the voices were used in informational and conversational in-vehicle systems. The next section contains a description of additional studies and how they conform to findings in literature as well as the findings presented in chapters 2 - 4.

5.2 Summary of Additional Studies

5.2.1 Conversational In-vehicle System

This study was conducted in collaboration with Oxford Brookes University, and parts of the results were published at IASTED Human Computer Interaction 2008 (IASTED-HCI 2008) (Jonsson 2008). I designed the study and the driving course. The questionnaires were either standard questionnaires, or the generic set of pre- and post-driving questionnaires designed by the group conducting driving simulator studies at Stanford University. The prompts for the in-vehicle system were designed in collaboration with Mary Zajicek at Oxford Brookes University in the UK.

In this study we investigated the effects of matching or mismatching the age-of-voice of a conversational in-vehicle system with young and older adult drivers. This investigation was motivated by the fact that car manufacturers (as well as insurance companies, and rental car companies) are starting to introduce more interactive and conversational systems to provide drivers with personalized services, and assess driver abilities. The in-vehicle information system is instrumented to learn about its driver(s), often by a) automatically collecting information about driving habits and ability, and b) interacting with driver(s) to learn about drivers’ preferences.

The study was designed as a follow-up study to the age-of-voice study presented in Chapter 4. In particular, we set out to investigate the effect of a conversational information system on driving behaviour and attitude when matching age of voice for the system with the age of the driver. To promote interaction with the system we designed the conversational system with questions and statements containing self-
disclosures. Jourard (Jourard 1926-1974) has shown that self-disclosures is an
efficient approach to extract information from people.

The experiment was conducted in a driving simulator, and the simulator was
instrumented to collect and save data on driving behaviour. The data included
different indicators of driving performance such as obedience to traffic laws (e.g.,
stopping at stop signs, not excessively exceeding the speed limit), ability to stay in
lane, brake behaviour, speed and accidents. Thirty six participants from two age
groups; eighteen between the ages of 55 and 75, and eighteen between the ages of 18
and 25, used a driving simulator in one of three scenarios; a conversational system
with a young voice (20 years old), a conversational system with an older voice (76
years old), and no system at all.

The results show that all drivers, regardless of age, were more willing to communicate
with a young voice than an old voice in the car. As in the study presented in chapter 4,
all drivers could identify the approximate age of the voices used by the in-vehicle
system. They placed one voice in the age group 18-25, and the other voice in the 55+
age group. The voices were also evaluated and found equal for other voice
characteristics such as emotional colouring, quality and clarity. Drivers’ willingness
to communicate had a detrimental effect on driving performance. This was especially
ture for young drivers. Even though they were aware of the fact that it negatively
impacted driving performance the younger drivers could not help but talk to the
system - even if it meant shifting focus from the driving task (Jonsson 2008). The data
also confirm results from the age-of-voice study in chapter 4. Older adults prefer a
younger voice; they were more willing to communicate with a young voice than an
older voice while driving. Once again the findings that contradicts similarity attraction
when older adults are sitting behind the steering wheel.

5.2.2 Blame Attribution and Drivers’ Attitude and Performance

This study was conducted in collaboration with Stanford University, California, USA
and Toyota InfoTechnology Center in Palo Alto, California, USA. Parts of the results
were published at Computer Human Interaction 2004 (CHI 2004) (Jonsson, Nass et
al. 2004). In this study we raised the question of blame attribution and self-serving
bias. The pre- and post driving questionnaires were designed by the group working
with driving simulator studies at Stanford University. I selected the questionnaires for
trust and liking and designed driving performance measures. The speech prompts for
the in-vehicle system were designed in collaboration with Clifford Nass.

The study investigated properties of in-vehicle systems and in particular how
linguistic cues, choice of words or phrasing of a message affect drivers. In this
experiment we studied drivers’ reaction to messages regarding their driving performance. In particular we investigated the effect of having a system comment about bad driving.

The study was based on blame attribution and self-serving bias (Miller and Ross 1975). Blame attribution describes how an individual moves from perceiving a situation to making a judgment of blame. Shaver (Shaver 1985) describes blame attribution in terms of cause, responsibility and blameworthiness. Even though people are most often associated with cause, it is possible that people attribute responsibility and even blame to the inanimate. In doing so they would either have to give the inanimate object some human qualities, or alternatively have it be under the control of some individual(s). Shaver also claims that intent (or lack thereof) can reduce blameworthiness (Shaver 1985). Self-serving bias (Miller and Ross 1975) explains why people attribute their successes to internal or personal factors but failures to situational factors beyond their control. This is typically used to explain why individuals tend to take credit for success but to deny responsibility for failure (Miller and Ross 1975).

The messages spoken by the in-vehicle system were phrased in three different ways based on blame attribution, blaming the driver, blaming the driver and car as a team, and blaming the driving environment. The messages were all warnings related to their driving performance.

The driver-blame system used sentences such as “You are driving too fast” and “You should slow down when taking these curves.” The we-blame system used the same sentences simply replacing “you” with “we,” viz., “We are driving too fast” and “We should slow down when taking these curves.” The environment-blame system would indicate external problems, and used sentences like “This road is easy to handle at slow speeds” and “These curves require slow speed.”

There were 36 drivers in the study, 18 males and 18 females that drove the simulator for two 20 minute sessions. Participants were randomly assigned one of the three versions of the in-vehicle system while driving. In addition to this, all participants also drove the driving course in silence. The order of the sessions was balanced so that half of the participants drove in silence before driving with a system, and the other half drive the sessions in reverse order.

Drivers’ attention to the road was measured using a “honking task.” Participants were asked to say the word “honk” as soon as they heard a honk. Fast response times to
honks were used to indicate high attention to the road and the driving task, while slow response times indicated low attention to the road.

Results show that phrasing of messages matter, and that blaming the environment worked best. Drivers felt most at-ease, they liked the system, and they rated the quality of the car higher. Their measured attention to the road was better than for drivers with the other systems, and most interesting also better than when drivers were driving in silence. These results conform to both blame attribution and self-serving bias. Drivers could experience the environment blame condition as a system where other drivers and objects in the driving environment could cause bad driving, most likely without the intent to harm. Any bad driving would hence be caused by someone or something else. The result was that the best liked system was the one that blames the environment, i.e. if I am driving badly it is somebody else’s fault.

5.2.3 Cognitive load and In-Vehicle Systems

In these two studies, we investigated how drivers’ cognitive work load affects their performance and attitude while driving with in-vehicle systems. A different in-vehicle system was used for each study. The first study was conducted in collaboration with Chalmers Technical University, Sweden, and the second study was conducted in collaboration with Oxford Brookes University, UK. Part of the results from the first study has been published at Human Computer Interaction International 2007 (Jonsson and Chen 2007). No results or material has yet been published from the second study.

For these two studies I designed the studies, the driving courses, the speech prompts for the in-vehicle system and the questionnaires. The exceptions are the general pre- and post-driving questionnaires designed by the group conducting driving simulator studies at Stanford University.

The first study focused on how an in-vehicle system that will give the driver information and traffic situation updates is perceived in low traffic and dense traffic situations. There were 17 participants in the study, 10 males and 7 females, and they all drove a driving simulator with an in-vehicle system for approximately 20 minutes.

The driving scenario was designed to have low density traffic situation and high density traffic situations. The in-vehicle information system gave information about the upcoming traffic situation for half of the traffic events, and was silent for the other half. This enabled us to investigate how information prompts from in-vehicle system was perceived and influenced driving performance in both low high density traffic situations, and also when no information was provided by the system. The study was a within-subjects design, so that all participants drove through all conditions of the study. The order of traffic density and information from the system
was balanced across conditions. This setup enables us to study driving behaviour and attitudinal responses in 4 combinations of system and traffic density:

- Responses in light traffic when assisted by the in-vehicle system
- Responses in heavy traffic when assisted by the in-vehicle system
- Responses in light traffic when driving without the system
- Responses in heavy traffic when driving without the system

Data was automatically gathered by the driving simulator for a comparison of how drivers react to high and low density traffic, and complex traffic situations, both with and without the in-vehicle system. Participants were also asked for their subjective evaluation of trust of the system and how they perceived it influenced their driving performance. Preliminary results show gender differences for both driving performance and attitude. More specifically, the data show that the in-vehicle information system was perceived to be more helpful in light traffic situations than in heavy traffic. This is confirmed by both measured driving performance data and attitudinal and perceived performance data from questionnaires. The data show that female drivers would benefit the most from this type system (confirms study on accuracy presented in chapter 3). There is a significant improvement in driving performance for female drivers when the system is in use. This is true when averaged over both light and heavy traffic. There was no positive effect of the system, averaged over light and heavy traffic, for male drivers. Female drivers also felt more calm and relaxed when driving with the system than when driving without a system. Male drivers showed no difference in state of mind driving with or without a system. The only effect for male drivers was improved driving performance when the system was used in light traffic.

The second study compared how an interrupting or non-interrupting in-vehicle system affects drivers. One system gave information when the driver was busy with a secondary task. The other system informed drivers when they were driving without being engaged with a secondary task. Two age groups, 18-25 and 55+, were recruited for the study to investigate if an interrupting system affects these age groups differently. Study participants were kept busy with arithmetic tasks at certain locations in the driving course. The in-vehicle system was designed to give hazard and warning information during the driving session; all messages were recorded in a young female voice. For half of the participants, the in-vehicle system interrupted with information during the arithmetic task, for the other half the system gave information either before or after the arithmetic task. There were 32 participants for
the study, 20 gender balanced drivers in the age group 18-25 and, 12 gender balanced drivers in the age group 55+.

The driving simulator was instrumented to automatically gather data during the driving sessions. In addition to this, participants were also asked to fill in questionnaires on their perception of the in-vehicle system and how they felt that the system influenced their driving performance.

In general, the data show that an interrupting system has a negative influence on driving performance, all drivers drove worse (more traffic violations and accidents) when driving with a system that interrupts them. Preliminary results suggest that young drivers felt more confident driving with the system than older drivers, and drivers with a non-interrupting system felt more comfortable than drivers with an interrupting system. Older drivers felt more helped by the system (both versions) than younger drivers, and all drivers felt that they reacted faster with a non-interrupting system. There are also age differences in how drivers felt the system influenced their driving. Older drivers felt safer with the interrupting system whereas younger drivers felt safer with the non-interrupting system. It is interesting to note that older drivers also perceived a greater benefit from the interrupting system than the non-interrupting system. This is contradicted by measured performance that clearly shows how interrupts negatively impacted their driving. Young drivers on the other hand, preferred the non-interrupting system with the perceived benefit visible in improved driving performance (fewer traffic violation and accidents).

Results also show that dialogues and conversational systems should be designed with care to minimise interruptions. Most deployed dialogue based systems require that the driver is stationary when engaging in dialogues. This seems prudent with current design since these systems are acting blindly. These dialogue systems have no knowledge of the driver’s workload, current driving environment or traffic. Dialogue systems in cars are still designed as primary task systems, whereas it is becoming clear that they are indeed secondary task applications. As such they should use information about drivers’ situation and traffic to engage drivers in dialogue or pause dialogues to ensure driving safety. Combining attractive voice selection with a secondary task design of the dialogue system, and in-vehicle system could then time the conversation to situations with light traffic where drivers can safely divide their attention between driving and talking. The results indicate that interruptions in-vehicle systems have a detrimental effect on performance. This is consistent with human to human dialogue where interruptions are considered socially unacceptable and also have a detrimental effect on performance.
Results from these studies are consistent with results from literature and previous studies. Both studies clearly show that drivers don’t like to be interrupted when they are busy, and their measured performance data confirm the negative impact on driving. This confirms a recent study by Mark, Gudith and Klocke (Mark, Gudith et al. 2008) that found that interrupts cause stress regardless of context. Even if people find the interrupt useful, it is still associated with a performance cost (Mark, Gudith et al. 2008). The results from the first study also confirm results from the study on accuracy presented in chapter 3. A female voice was used in this study, same as in the accuracy-study (chapter 3). Female drivers tend to benefit more from the system even if it is interrupting drivers while concentrating on complex traffic situations. Results from the second study confirm results from the age-of-voice study presented in chapter 4. Older adult drivers like and appreciate the system more than young drivers. They appreciate the information regardless of if the system interrupts them or not, and they are not aware of the negative effect interrupts has on their driving performance.

5.2.4 Matching Personality of Driver with Personality of Voice

This study was run in two locations, at Oxford Brookes University in the UK and at Stanford University in the USA. There were 40 drivers in the UK study (20 Dominant and 20 Submissive) and there were 60 participants in the US study (20 Dominant, 20 Submissive and 20 Middle – neither Submissive nor Dominant). No results or material has yet been published elsewhere from these two studies. I designed the study for the UK, and co-designed the study for the US with Clifford Nass. I designed the driving course and the driving questionnaires, except the general pre- and post-driving questionnaire designed by the group conducting driving simulator studies at Stanford University. The personality questionnaire is a standard commercial questionnaire. The speech prompts for the in-vehicle navigation system was designed together with students from Stanford university and Mary Zajicek from Oxford Brookes University.

In this study on how personality of voice affects driving, linguistic and paralinguistic cues were combined to indicate characteristics of personality. Previous studies (Nass and Lee 2000) on matching personality of computer program and users have showed that similarity attraction holds. A system is always rated higher, and the user’s perceived performance better in matched cases, where the system has the same personality as the user. We wanted to investigate if this was true for an in-vehicle system.

Participants for the study were assessed on the dominant-submissive dimension using the Interpersonal Adjective Scale (Wiggins 1995) This dimension represents the degree to which an individual is assertive and willing to exercise control over others.
An in-vehicle navigation system was designed so that, in addition to giving the driver directions, it also gave the driver choices at certain points along the route. There were two versions of the navigation system. One version used a dominant male voice for navigation information, and the other version used a submissive male voice. Even though the information was the same in both systems, down to the numbers of words, the utterances varied in choice of words and voice characteristics. The dominant voice used words such as “will”, “must” and “definitely,” and the submissive voice used words such as “might”, “could” and “perhaps”. When the dominant navigation system used assertive language “You should definitely turn right” the submissive system was more timid, “Perhaps you should turn right”. For choices the dominant voice would, for instance, say “Continuing straight is shorter but may have more traffic. Turning right will definitely be faster” and the submissive voice would say “Continuing straight is shorter but may have more traffic. Turning right will probably be faster”. The voices also varied to represent dominant and submissive traits. The dominant voice was given a higher overall frequency, a larger range of pitch during speech, and greater speed than the submissive voice (Nass and Lee 2000).

Results suggest that 1) words and intonation matter, and 2) drivers can discern the personality of the voice used by the in-vehicle system. Most important however, results from the US study do not confirm findings based on similarity attraction.

5.2.4.1 Results from US study
The navigation system with the dominant voice was preferred by all drivers, both dominant and submissive. Data show that all drivers, regardless of personality, followed directions when given by the navigation system with the dominant voice. There was less attention paid to instructions given by the navigation system with the submissive voice, even for submissive drivers. Data also suggest that even though there is a dominant driver personality bias for more accidents and breaking of traffic rules, personality of car voice can influence drivers. Drivers with the submissive voice, regardless of personality, tend to: 1) drive slower and heed speed limits more than drivers with the dominant voice, 2) break fewer traffic rules than drivers with the dominant voice, but 3) have more accidents (off-road accidents and collisions) than drivers with the dominant voice.

For the “middles” (neither dominant nor submissive) the system with the dominant voice was preferred over the system with the submissive voice. Drivers with the dominant voice followed directions by the systems better than drivers with the submissive voice. The “middles” also show the same trend as the dominant and submissive drivers, when driving with the navigation system with the submissive voice, with slower speed, fewer traffic violations but more accidents. The preference
for the dominant voice as expressed by the “middles” (seen in attitudinal measures) confirm results in a study on personality by Nass and Lee (2001; 2003). Their results showed that using a dominant voice was preferred when the personality of the listener was unknown. The anomaly for measured performance (driving behaviour) found in this study however, warrants further investigation.

5.2.4.2 Results from UK study
Data from the UK study differs slightly from the US study. Even though the navigation system with the dominant voice was preferred by all drivers, and that everyone followed directions better when given by the dominant voice, the data show better alignment with similarity attraction. There is a personality bias so that dominant drivers have more accidents and break more traffic rules. Nonetheless, the data also show that mismatched conditions perform worse than matched conditions. Dominant drivers that drive with submissive voice (and submissive drivers that drive with the dominant voice) have more off-road accidents, exceed the speed limit more often, run more traffic lights and miss more stop-signs, than matched conditions. Overall, the submissive voice seems to have less negative impact on UK drivers than on US drivers. I will discuss more on this topic under the section on cultural differences later in this chapter.

It is interesting to note that there is a mismatch between perceived performance and measured performance. All drivers preferred the dominant voice and they also perceived that they drove better with the dominant voice; this contradicts earlier findings in similarity attraction. Analyzing measured performance, voice has a very strong influence on driving behaviour. US drivers show a mixed reaction to the submissive voice, they drive slower and break fewer traffic regulations but have more accidents. UK drivers show reactions consistent with similarity attraction, better performance for matched conditions. This mismatch in perceived and measured performance warrants further investigation.

5.2.5 Drivers’ Cognitive styles and In-Vehicle systems
My latest studies have focused on how the perception and influence of a speech based in-vehicle system would interact with drivers’ cognitive abilities and cognitive styles. We have two studies on this topic, one at Oxford Brookes University in the UK, and the other at Linköping University in Sweden. No results or material from these studies have yet been published. The studies were designed in collaboration with Nils Dahlbäck, Linköping University. The questionnaires for cognitive styles (visual or verbal) were selected in collaboration with Nils Dahlbäck from Linköping University. The driving courses and all other questionnaires were re-used from the personality study.
In addition to cognitive style, results suggest differences in how native and non-native speakers are affected by both selection of language and gender of voice in an in-vehicle system. Non-native speakers have difficulty hearing and understanding language with even small disturbances, such as added noise and missing syllables. These disturbances are similar to the effect that a noisy car environment has on the messages spoken by an in-vehicle system. This negative trend is further exacerbated by the gender of voice on the in-vehicle system. Female voices have been found to carry better in noisy environments (Nixon, Anderson et al. 1998; Nixon, Morris et al. 1998; Edworthy, Hellier et al. 2003). Preliminary results from this study seem to confirm this for the vehicle environment. More importantly however, is that the results indicate that this effect is a magnitude larger for non-native speakers. The use of a male voice in the in-vehicle system seems to render the system almost useless (unintelligible and disliked), whereas the female voice is both better understood and liked.

5.3 Emerging Patterns and General Observations

An emerging pattern from the studies presented here indicates that attitudinal measures are linked to behavioural measures. In general, drivers that are pleased with and trust the in-vehicle information system also show better performance than drivers that dislike the system. This is consistent with the literature that positive affect improves performance and problem-solving (Isen, Daubman et al. 1987; Groeger 2000; Isen 2000). There are however also findings from some of the studies presented here that contradict this and show that drivers can like and trust an in-vehicle system and believe that it makes them better drivers even though measured performance data show the opposite. For example, the personality study (section 5.2.4), shows a preference for a dominant voice, even though the performance data show that groups of drivers drove better with a submissive voice. Data from the conversational study (section 5.2.1), shows a similar trend, even though the in-vehicle system with a young voice was liked and trusted by all drivers, the performance data show that they drove worse with more incidents than with an older voice.

Findings presented here also show that even though, in general, measured performances follow perceived performance, data from some studies contradict this. For example, data from the study of the interrupting in-vehicle system (section 5.2.3) show that older adult drivers perceived the interrupting system to be helpful and aid driving performance, whereas in reality, measured performance data show that the interrupting system had a negative impact on. Data from the personality study (section 5.2.4) also show similar trends, drivers followed directions from a dominant voice and perceived the voice to have a positive influence on their driving.
Even though it was generally true that there were positive correlations between attitudinal measures and performance measures as well as between perceived and measured performance, the contradictions highlighted above give rise to some questions. Can we generalize to liked and trusted automatically means better performance? Can we accurately verbalize the factors that form the base for actions and decisions? Further study is required to answer these questions.

Another emerging pattern is that similarity attraction does not necessarily hold in the car. Data from the age-of-voice study (chapter 4) show that older adult like and trust the older adult voice in the office but not in the car where they prefer the young voice. Media studies show that similarity attraction and homophily are important and solid findings, people tend to like and trust people that are like themselves. Studies from SRCT at the Department of Communication, Stanford University, show that similarity attraction holds for personality studies in office settings. Findings from the personality study presented in this thesis (section 5.2.4), on the other hand, show that the dominant voice is preferred by all drivers regardless of personality. These results clearly do not confirm similarity attraction, so why? What makes the driving environment different?

A possible explanation for the car being different is linked to real time operating environments. In the car the driver is an operator, controlling a vehicle speeding through space and time. In the office, the individual sitting by the desk is clicking on links and filling in questionnaires. There can of course be consequences of badly planned actions in both settings. It is possible that the differences found in the data presented here are based on the difference in immediate feedback and the potential threat to life and limb? This is not to belittle consequences of bad investments or less fortunate online choices, but instead trying to set the scene for an operator controlling an object where mistakes can have immediate life threatening consequences. It is after all in this setting that the data from the studies presented here deviate from studies in social sciences.

An alternative perspective, from social cognition and cognitive psychology, is that the driving environment is a situation where the theory of dual-processing accounts for behaviour (Evans 2008). In this theory there is a distinction between fast and automatic cognition and slow and deliberate cognition and that this influences behaviour. Norman (Norman 1993) makes a distinction between experiential and reflective cognition, where experiential cognition is associated with expert/skilled and reactive behaviour, and reflective cognition with slow/planned and reconsidered behaviour. Driving is an activity done with seeming ease and skill by the experienced driver. The driver continually takes in multiple information streams, through the
Summary, Discussion and Conclusions

Windshield, from instruments, tactile feedback from seat, steering wheel and pedals, and the sound of the engine. The driver’s responses seem effortless and without delay. The same behaviour is true for all domain experts; even though enormous amounts of processing are necessary the responses are generated without conscious effort. Evans’s review of proposals (Evans 2008) for dual processing in cognition found that in a study of fire officers and paramedics, experts recognized situations encountered previously and rapidly made decisions. The quality of this decision making is affected by cognitive capacity and concurrent working memory load. Norman phrases this as “being experiential or reactive”, driven by the patterns of information arriving at our senses, but dependent upon a large reservoir of experience (Norman 1993). Deliberate cognition or reflective thought is different; it engages problem solving, planning and assessing alternatives before decisions are made and actions taken. Even though driving is most often a data-driven reactive activity in need of quick decisions, there are complex situations where reflective planning and assessment is necessary.

Research on in-vehicle information and support systems takes much of its cues from research in aviation. In the same manner that results from aviation are applicable to vehicles, it would be reasonable to assume that results from research on in-vehicle systems can be applied to other real time environments. Using speech messages in these kind environments opens up questions about sensitivity to errors. This sensitivity must be more pronounced in environments where operators rely more on instruments than on the five senses. The response to errors, however, is probably the same; users start to ignore the system if the error rate is too high. This would increase the threshold for when a warning or information messages is deemed necessary to issue. The result would be to omit information in favour of keeping a high accuracy level.

Data from the conversational study (section 5.2.1) also show that it is not always the best liked voice that leads to the safest driving practices. There are tradeoffs in the selection of voice, the attractiveness of a voice should be balanced with the effect it has on driving safety.

Furthermore, the halo effect found in the studies described in chapters 2 to 4 is emphasized in all additional studies. The perception of the voice influences not only the perception of the in-vehicle system but also the perception of the car. Automobile and in-vehicle system manufacturers should select voices carefully to ensure positive association with brand.
5.4 Limitations and Methods

5.4.1 Cultural, Age and Gender Differences

I have mainly studied in-vehicle systems in the US focused on the age group of 18 to 25, with a couple of studies in the UK comparing two age groups, 18 to 25, and 55+, and a couple of studies in Sweden. All participants in the younger age group, in the US, the UK and Sweden, were recruited from universities. The older age group in the UK was recruited from an affluent socioeconomic group. All studies have been gender balanced across conditions.

It is clear, even from the limited number of studies that I have done with drivers from different countries that there are cultural differences. This is seen both in driving habits and how drivers react to voices and voice systems in cars. A general observation is that US drivers typically spend more time in their cars than UK and Swedish drivers. This is especially true for drivers from California where public transport is scarce. UK drivers in the Oxford area on the other hand, are encouraged to use public transport such as busses and trains, as are Swedish drivers in Gothenburg and Linköping. For the younger group of drivers, I have found very few differences between US and Europe, mainly just nuances in verbosity and reaction to submissive voices. The conversational study, reported in section 5.2.1, was conducted in the UK and shows that participants in the experiment were all willing to engage in a dialogue with the in-vehicle system. Overall, however, results show that US drivers tend to engage more in dialogues with the in-vehicle systems than drivers from Europe. The personality study, conducted both in the UK and the US, show that even though the dominant voice personality works better than the submissive voice personality, the bias is less in the UK than in the US. It is hence not even safe to assume that results can be transferred between English speaking countries. The differences found in the personality study are important and indicates that the design process of in-vehicle systems should always include testing in target countries.

In the studies where interactions between age of voice and age of driver have been investigated, results clearly show age differences in attitude and driving performance. There is bias towards a young voice in the car regardless of driver age, a result contradicting homophily and similarity attraction. A possible explanation would be that the car is seen as a real time operating environment and the young voice is associated with good eyesight, and quick reactions. There is also a clear difference in driving performance data between older and younger drivers. In general, older drivers drive slower (fewer speeding tickets) and with fewer incidents. A possible explanation for this difference would be in lower propensity for risk-taking by older drivers.
Summary, Discussion and Conclusions

There are many nuances of gender differences emerging from my studies. First, literature shows a positive bias towards female voices in noisy environments (Nixon, Anderson et al. 1998; Nixon, Morris et al. 1998; Edworthy, Hellier et al. 2003). In one of my studies, we found indications that female voices work better for non-native speakers than male voices. The female voice was perceived to be more intelligible than the male voice in the car by the non-native speakers. Most non-native speakers did not comprehend the navigation instructions when presented by the male voice, whereas the same instructions presented by a female voice were understood. This result is an indication that gender of voice, especially for rental cars or cars where non-native speakers are likely to be drivers, should be further investigated.

Second, results from all studies in the age group of 18-25 show a clear and consistent gender difference for driving performance. Female drivers perform consistently better than male drivers; they have fewer accidents and adhere better to traffic regulations such as speed and stopping at stop signs and traffic lights. This confirms statistics from insurance companies (visible in premiums for car insurance in the US). US data show that male drivers are 2.9 times more likely to have a fatal accident than female drivers. Based on miles driven, male drivers in the United States are 77% more likely to be killed in a car accident than female drivers. This difference is higher for drivers 18 to 25 where the data show that male drivers are more than 3 times more likely to have a fatal accident than female drivers. This difference is also present, although not as significant, in driving performance in age group 55+. Male drivers in this age span are 2.5 times more likely to have a fatal accident than female drivers (see http://www.traffic-stats.us/ for US statistics). Data also show that 82-year-old women belong to the highest risk category closely followed by young male drivers aged 16-25. However, this is a function of frailty rather than safety, since the same group of older women is equally more likely to die from any accidental cause.

Statistics also show that men get more speeding tickets than women do and the amount over the posted speed limit is greater for men than women. Higher speeds mean more damage to both people and property if an accident occurs. Even though men and women make about the same number of claims a year but women's claims are usually for small bumps or prangs while men make bigger claims and make most of the claims for writing off their car. This inequality is not visible in premiums where young female driver are still expected to pay the same high premiums as young male drivers. The same inequality is seen in age group 55+ from the UK that shows that UK's safest driver is a woman aged 55. She is however rewarded with lower premiums (~10%)
This performance difference was also found in the simulator studies with higher differences between genders in the age group 18 to 25 and smaller performance differences (if any) for drivers 55 years of age or older. Even though the data from the simulator show the same gender bias for performance as in real driving, we investigated potential sources for this bias. One potential reason could be that drivers in the younger age group were influenced by using video games. According to questionnaire data collected from the studies questions about the use of video games indicate that there are differences between age groups but not between genders. The older age group did not play video games at all and the younger group played frequently. There was no significant difference in how much time male and female participants spent with video games every week.

Another potential reason for this gender bias for performance, as pointed out by Clifford Nass, Stanford University, is that female participants always tend to focus more on the task of a study than male participants with a consistent result of better performance. This is confirmed by Tannen (Tannen 1990; 1995) her studies show that women tend to be inclined to do what is asked of them, while men tend to resist even the slightest hint that anyone, especially a woman, is telling them what to do.

Third, there are potentially gender differences (based on studies in the US so the results could potentially be cultural biased) related to the gender of voice of the in-vehicle system. Even though both genders seem to have negative stereotypes about women when it comes to issues of driving (Nass and Brave 2005), women show a positive bias toward females voices (Nass and Brave 2005). If men dismiss a female voice, they would be less likely to be influenced by the messages from and interactions with the in-vehicle system. It is interesting to note that most deployed navigation systems use female voices; most likely based on data from the aviation industry that show that a female voice carries better than male voice in the cockpit (Nixon, Anderson et al. 1998; Nixon, Morris et al. 1998; Edworthy, Hellier et al. 2003). BMW hence selected a female voice for their in-vehicle information system for one of their high-end models. This turned out to be a bad choice. The demographics of the buyer for that model did not react well to a female voice in the car. After recall, re-design and re-release with a male voice for the in-vehicle system, the model was successfully accepted by the target group. This is an indication that voice selection for in-vehicle systems need to be thoroughly tested in context with the correct user group before deployment.

An alternative explanation to the reaction of gender of voice is found in attention to information. This theory, based on a series of studies – once again conducted in the US with a potential cultural bias – by Meyers-Levy (1988; 1989) demonstrates that
women feel more responsible for attending to information from other sources. Women try to take all information into account, even details, and men try to ignore detailed information flows to focus on broad message themes. Women attempt to understand and incorporate all available information before rendering judgment. Men are more likely to tune out information. This might explain why women tend to perform better with in-vehicle systems than men. This is supported by Tannen’s (1990) extensive work on gender and communication provides a different explanation, once again studies done in the US with a potential cultural bias. Tannen’s studies show that women use speech to establishing relationships, while men use speech to establish status and a hierarchy of superiority. Men are more comfortable giving advice than accepting advice; women are equally comfortable giving and receiving information and advice. This could make men perceive information and advice from an in-vehicle information system as an attempt to adopt a position of hierarchical superiority, while women would welcome the information perceiving that the system is trying to be helpful. (See section 3.6.1). It is relatively straightforward to investigate the effect of gender of voice on driving performance by conducting studies where in-vehicle systems include conditions for both male and female voices. The other alternative explanations could not be as easily tested. Nonetheless, there are ample results of gender bias towards voices (Nass and Brave 2005).

Please note that age, cultural and gender studies are research topics in their own right, and further studies should be done to refine results presented above.

5.4.2 Driving Simulators and Measures
Many criticisms of driving simulator-based studied centre around the fact that the data produced by simulators lends itself to analysis of raw numbers of accidents and traffic violations rather than a more complicated examination of fine grained driving behaviour. While it may seem logical to focus on the nuances of steering, lane position, acceleration and braking patterns, this information is not readily available for real drivers. In fact, the number of accidents that drivers have and the number of moving violations may seem crude, but these figures are exactly the numbers that private and public agencies use to evaluate drivers. The Department of Motor Vehicles in US states uses a point system based on speeding and accidents, among other things, to judge whether or not drivers should keep their licenses. Similarly, insurance companies use these numbers to determine premiums. While it would be nice to have a thorough understanding of all nuances of driving behaviour, especially in response to speech based interactions with in-vehicle systems, it is simply not feasible, either in a real-life or simulated scenarios. What is important to understand is that the crude numbers of driving accidents and violations serve the same function in both real and simulated driving. These numbers indicate critical breakdowns in driver
attention, judgment, and vehicle management; it is these failings that predict future driving problems. Thus, when looking at behaviour in driving simulators, as well as in the real world, we look to accidents, speeding, and traffic violations to give us valuable insight into patterns of driving behaviour.

5.4.3 Driving Simulators and Fidelity

There are many different types of driving simulators, ranging from simple lab settings to vehicle based driving simulators with fixed or moving bases. Different methods and tools are furthermore used to assess driving performance, driver status and driver interactions. This leads to questions about replication of studies, comparisons of results and validity of results.

The use of driving simulators for research and other purposes is becoming more widespread. As a direct result of widespread use, the availability of different types of simulators has increased at the same rate that the price has decreased. Hence, there is now a wide variety of different driving simulators with varying properties, fidelity and levels of configuration. Starting with game-like driving simulators where “drivers” use common game accessories to control and drive scenarios displayed on computer screens/televisions (Electronics Art Inc 1998; Polyphony Digital Inc 1998), to fixed-base simulators (Systems Technologies Inc 1990; DriveSafety Inc 1995) and moving-base simulators (VTI 1984; National Advanced Driving Simulator 2002).

The ultimate goal of simulators and simulator studies is to mirror all aspects of driver behaviour and performance (Green 2005). Given the range of simulator types, it is however hard to validate results from simulator studies without a detailed description of each simulators limitations and abilities. Results from driving simulator studies concentrate on the driver or driving functions that are studied rather than on properties of the driving simulator setup (Green 2005). To enable validation and replication of driving simulator studies, it is however important to describe all aspects of the study, including participant selection, technology, configuration, methods and measures (Lee 2004).

Technological advances have led to simulators with increasing levels of physical fidelity with the assumption that higher levels of fidelity will produce better results and thus more accurate insights into driver performance. Lee (2004), however, claims that low-fidelity simulators may be more effective than those that aim for a visually and behaviourally realistic representation of the driving environment and vehicle dynamics. Low fidelity simulators or simulators that intentionally distort the driving experience have the potential to effectively target certain aspects of driving better than high fidelity simulators that strive for a life-like representation of the driving
environment and vehicle dynamics. This has important implications for selection of simulator type, scenario design, scene rendering, and motion cues. Identifying the lowest possible fidelity for a given training objective, assessment protocol, or experimental question forces a careful evaluation of driving that might not otherwise occur.

Driving simulator studies are often setup with tailor-made scenarios to investigate driving behaviours related to, traffic events, use of one or more in-vehicle information systems, or driver impairments. There is no standard for the description or design of the driving scenario. Measures used in driving simulator studies are relatively well known and cover three groups of driver behaviours fairly well, behaviour of the driver, behaviour of the car, and attitudes towards in-vehicle systems and car. Commonly used measures include neuropsychological tests, performance of simulated vehicles, reaction time measures, and assessment of visual abilities (including acuity, night vision, and depth perception). It is hard validate results and replicate studies since there are no standard model or guidelines for which scenarios and measures to use and when (Jonsson 2006). There are however ongoing efforts to standardize driving simulator studies (for example by Rizzo and Lee from University of Iowa – see http://www.simusers.com (Simusers-Wiki 2009)).

Östlund, Carsten et al. (2004) report that the fidelity of the simulator seems to influence the results very little. The authors conducted a meta-study where one of the objectives was to identify the advantages and disadvantages of the different assessment methods including laboratory, simulator, and field. There were 17 experiments in total with 527 participants that investigated two in-vehicle system tasks in three different roadway settings in various European countries and Canada. There were some inconsistencies in the results from the studies by Engström, Johansson et al. (2005) and Jameson and Merat (2005). Drivers increased speed while interacting with an in-car system as reported by Engström, Johansson et. al. (2005) while drivers in the Jameson and Merat (2005) study slowed down. This difference was attributed to differences in driving scenarios and participant groups when analyzed in the meta-study. The Engström, Johansson et. al. (2005) study also show larger variations in lane behaviour in the fixed based simulator than in the moving-base simulator. The meta-study however showed no significant differences in results obtained from different fidelity simulators, ranging from lab based setups to moving-base simulators (Östlund, Carsten et al. 2004). The field studies on the other hand tended to pick up somewhat different effects of the in-car systems than the simulator studies. Some effects are based on haptic/visual cues – lane behaviour and distance – and other effects are based on tunnel vision that affected attention to peripheral
objects in the field. Simulator and field studies can hence be seen as complementary (Östlund, Carsten et al. 2004).

5.4.4 Driving Simulators and Driving Cars

A review of driving simulator studies and results from these studies show that some results are transferable to reality. In a study by Engström, Johansson et al. (2005) the authors investigated if results from driving simulators and from driving an instrumented vehicle on the road were comparable. Their findings indicate strong similarities in driving behaviour between driving simulators and driving on a real road. Hoffman, Lee et al (2002) also report that findings from a study where drivers’ braking responses were investigated both in a high fidelity simulator and on a test track transfer. The results show that the advantages for simulator studies include more and precise control of lead vehicle and headway to brake onset of the lead vehicle. Disadvantages for simulator studies include increased risk of simulator sickness with repeated braking and imperfect visual cues. Overall, results show that the general patterns of driving behaviour are similar in most respects between simulator and field, where minor differences in results between driving simulators and real driving could be explained by visual and haptic cues.

What seems to hold for driving simulators is that they can be successfully used for training purposes and screening. For driver training, simulators seem to provide real-life experience enough to produce positive results. For screening, driving simulator studies of new in-vehicle systems seem useful for identifying systems and features that need more work before they should be tested in real cars.

It is interesting to note, based on discussion with the head of the Toyota Formula One race team, that driving simulators are not particularly useful for race car drivers. Regardless of the fidelity (based on tests in Toyota’s simulator) the generated visual and haptic cues were not real enough to satisfy these drivers. This is contrary to data from the aviation industry where results show that pilots exhibit the same reactions in the simulator as they do in the cockpit. The difference here might be explained by the move to fly-by-wire for more advanced planes so that flying by the-seat-of-the-pants no longer is advantageous. This is not true for F1 racing, where the tradition still is to drive by the-seat-of-the-pants. When these drivers test cars equipped with drive-by-wire features, they simply turn them off to stop them interfering with their driving.

In summary, findings from comparisons of studies performed in different type driving environments indicate that even though the results are basically the same, there are small differences. This makes it important to perform testing of proposed in-vehicle systems in a real car in real traffic before deployment.
5.4.5 Generalizing Results from Studies Presented in this Thesis

All of my studies have been done using driving simulators, starting with “video-games” and then moving to commercial driving simulators. The most glaring deficiency when using a video game (blame attribution study reported in section 5.2.2 and emotional drivers reported in chapter 2) is that speech generated by the in-vehicle system was randomized and not based on participants’ driving behaviour, or location in the driving course. This problem was resolved in later studies where a commercial driving simulator was used. The commercial simulator allowed interactions initiated by the in-vehicle system to be scripted based on driving behaviour, road conditions and the driver’s current situation.

I have also selected to use driving behaviour measures such as accidents and failure to adhere to traffic regulations to measure driving behaviour. These measures are definitely less fine grained than the conventional brake behaviour and lane deviation measures used in most test track studies. Nonetheless, driving, in simulated as well as in real traffic, is a complex activity that continually tests drivers’ abilities to react to the actions of other drivers, traffic and weather conditions, and unexpected obstacles. Therefore, I chose to challenge people with hard and complex driving scenarios. In this way, the participants were subjected to a higher rate of potentially risky situations than they would be within a lifetime of driving. Mishaps and accidents in the driving simulator (as in real life) are signs of critical breakdowns in driver attention, judgement, and vehicle management. These driver mishaps and accidents were correlated with the use (or lack of use) of in-vehicle systems as measures of how these systems influence driving behaviour.

Can the results from my studies of in-vehicle information systems conducted in driving simulators be applied to cars? Based on the review in section 5.4.4, I would argue that the driving simulator can be use as an effective development filter, and that in-vehicle systems with a favourable impression on drivers and driving performance in the simulator should be tested in real cars to see if they are equally beneficial in real traffic. I would also argue that lapses in driver attention, judgement and vehicle control visible in driving simulator studies due to the influence of in-vehicle systems, would be present also in real driving. These lapses will hopefully not cause an accident in real traffic as it may have done in the simulator, but there will definitely be less severe expressions of bad driving. Based on this, it is my belief that lapses in safe driving behaviours and perception of in-vehicle systems and cars do transfer from driving simulator studies to real cars in real traffic.
5.4.6 Long Term Effects

For the results from these studies, it is hard to say how in-vehicle systems tested in a driving simulator for approximately 30 minutes would hold up over time. Will the benefits found from drivers interacting with a new type system wear off over time, or will the perceived and measured value of the system stay intact? There is always the potential that a system that seems friendly and informative when driving for 30 minutes, might be perceived to be too verbose after 2 weeks. There are opposing forces working to change behaviour over time, there is the chance that prolonged exposure would make a system both less frustrating and less exciting. It is also possible that even while novelty and wow-effects are crucial for initial use, properties including perceived usefulness and social capital might explain longer term use (Karapanos, John et al. 2009). Karapanos et. al. (2009) claim that there are three main forces to explain long-term adaption, familiarity, function dependency and emotional attachment.

Results from studies presented here conform to this model; data show that (positively associated) familiar voices when used by in-vehicle systems have a positive effect on upset and frustrated drivers. Familiarity of voices ties back into how long-term use of in-vehicle information systems affects drivers’ attitude and performance. To select voices for in-vehicle systems carefully, and then rely on familiarity to build over time, would ensure that positive effects would hold over time. This of course hinges on the assumption that no hidden or introduced negative properties of the system emerge. For instance, a voice that is trustworthy and liked can through unforeseen and/or external events become disliked and distrusted. For example, the voice of a well-known public individual may be used, and this individual may be caught doing something illegal, or simply stating an objectable opinion.

It is also possible that the long-term effect of a voice is different for different types of in-vehicle systems. A hazard and warning system is probably associated with a different set of selection criteria than a conversational system recommending shopping and leisure activities. Furthermore, it seems natural that a system that is engaging in a one-to-one communication, such as an in-vehicle system, would know the driver as well as the driver learns to know the in-vehicle voice. It would seem desirable to build a system where the in-vehicle system learns about its driver(s) to further increase trust and in reciprocity increased driving performance.

An in-vehicle system might also be perceived to be helpful when driving alone, but disturbing and socially unacceptable volunteering streams of information when the driver is conversing with passengers or interacting with other devices. My results
Summary, Discussion and Conclusions

show that in-vehicle systems need to be sensitive to the driver status and mood and also to the surrounding driving environment. Upset drivers benefit from voices that are downbeat. Drivers under high cognitive load simply don’t want information from an in-vehicle system. And drivers don’t like to be interrupted by an in-vehicle system while talking, see section 5.2.3. Moreover, the effect of communication with an in-vehicle systems while driving with passengers or interacting with other devices needs further investigation. What type systems can interrupt the driver? What type systems should cease interactions when passengers are present or when the driver is busy with another secondary task?

5.5 Summary and Final Comments

The studies presented in this thesis were designed to address the question:

Do voices matter in the car?

The results show that voices used by in-vehicle systems really do matter!

Voices, and more specifically characteristics of voices, used by an in-vehicle information system directly affect a driver’s attitude and performance. Previous studies on the effects of voice have been done in office or home settings. Voices trigger cultural and social affects that influence trust and liking. The work presented here affirms that voices are not neutral and trigger the same effects when used in the car, however it also shows that the car is different to the office setting.

Voices carry social cues. This leads individuals to react differently to the same voice. Therefore a single voice is not effective for all drivers. Results presented here show that for both attitude and driving performance, drivers clearly benefit from an in-vehicle information system that knows its driver. This suggests that voices used in in-vehicle systems should adapt to their drivers. Unfortunately, current technology is insufficient to fully implement these findings.

All studies presented in this thesis were done using driving simulators. However, while initial testing can be done in a driving simulator, I must emphasize that new in-vehicle systems should ultimately be tested using real vehicles in real traffic.

Even though all studies were done in a simulator, the immersive nature of driving simulators enabled results that show that the car environment is different than an office or home setting. My results furthermore show that social reactions to voice communication in the car are different than social reactions in the office environment. Importantly, the results show that the property of similarity attraction, an otherwise
solid finding in social science, does not necessarily hold in the car. Voices are perceived differently and affect people differently in the car than in the home or office environment. A possible explanation of this comes from the theory of dual-processing; the car is a real time operating environment requiring reactive decision making. Similar results would most likely also be found in other environments that rely on quick data-driven decisions.

Voices need to be selected so that people are willing to listen to them or communicate with them. Results presented here, however, also indicate that there is a trade-off between liking and performance. It is important to select voices that are attractive enough that drivers listen, but not so attractive that drivers talk instead of driving.

Taken together the studies presented in this thesis show that both attitude and performance can be improved by selecting an appropriate voice. It is not possible in this thesis to cover every aspect of voice impact, but this thesis does attempt to highlight the effects and importance of a number of voice related factors. Much more work is needed, investigating voice characteristics used in cars and other real time operating environments to complete the picture.

For a system to know when to talk and when to be quiet, the system needs the technology and know-how to monitor and adapt to both operator and environment. If a system can be designed to do this, it will most likely be safe to use by operators of any vehicle, be it a car, a ship or a plane. With clever and strategic references to all of the different rationales for behaviour change, cars and other technologies might dramatically enhance safety while encouraging positive feelings. What more could a designer want?
6 References


References


Dahlbäck, N., S. Swamy, et al. (2001). Spoken Interaction with Computers in a Native or Non-native Language - Same or Different? INTERACT, Tokyo, Japan.


Entemann, V. and H. Kuting (2000). Safety deficiencies of elderly driver and options provided by additional digital map content. ITS World Congress, Italy.


Guerrier, J. and S. Fu (2002). Elder Roadway User Program Test Sections and Effectiveness Study Talahassee, FL, Florida Department of Transportation Research Center.


Hakamies-Blomqvist, L. and P. Henriksson (1999). "Cohort effects in older drivers' accident type distribution: are older drivers as old as they used to be?" Transportation Research Part F: Traffic Psychology and Behaviour 2: 131-138.


References


Social and Emotional Characteristics of Speech-based In-Vehicle Information Systems: Impact on Attitude and Driving Behaviour

Factors in Driver Assessment, Training, and Vehicle Design, Rockport, Maine.


Klein, J., Y. Moon, et al. (2002). "This computer responds to user frustration: Theory, design, and results " Interacting with Computers 14(2): 119 - 140.


Niedenthal, P. M., M. B. Setterlund, et al. (1994). Emotional organization of perceptual memory. The heart's eye: Emotional influences in perception and


References


7 Appendices

7.1 Appendix A

7.1.1 Speech Prompts for Angry and Frustrated Drivers

The speech prompts in the table below was used for the study with induced angry and frustrated drivers. The driving course was divided into four equal segments and each segment contained the same number and type of prompts.

<table>
<thead>
<tr>
<th>Prompt</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>This road is very busy during rush hour, and traffic is slow.</td>
<td>Acknowledgement</td>
</tr>
<tr>
<td>I get stressed in traffic almost every day, how often are you</td>
<td>Management</td>
</tr>
<tr>
<td>stressed by traffic problems?</td>
<td></td>
</tr>
<tr>
<td>The stop sign in this intersection slows down traffic.</td>
<td>Acknowledgement</td>
</tr>
<tr>
<td>What do you think about traffic?</td>
<td>Management</td>
</tr>
<tr>
<td>This road is too narrow and windy for bicyclists.</td>
<td>Acknowledgement</td>
</tr>
<tr>
<td>What do you think about the driving conditions?</td>
<td>Management</td>
</tr>
<tr>
<td>The police often use radar here, so make sure to keep the speed</td>
<td>Suggestion</td>
</tr>
<tr>
<td>limit.</td>
<td></td>
</tr>
<tr>
<td>Do you generally like to drive at, below or above speed limit?</td>
<td>Management</td>
</tr>
</tbody>
</table>

**Part one (1/4 of driving course) had 3 Acknowledgement prompts, 4 Conversational prompts, and 1 Suggestion prompt**

<table>
<thead>
<tr>
<th>Prompt</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>How do you think that the car is performing?</td>
<td>Management</td>
</tr>
<tr>
<td>This is a really windy stretch of road.</td>
<td>Acknowledgement</td>
</tr>
<tr>
<td>What types of situations makes you feel stressed while driving?</td>
<td>Management</td>
</tr>
<tr>
<td>Traffic is really heavy, if you turn left you might avoid some of it</td>
<td>Suggestion</td>
</tr>
<tr>
<td>Pedestrians cross the road without looking in this school zone.</td>
<td>Acknowledgement</td>
</tr>
<tr>
<td>I like to drive with people that talk to me, what is your favourite</td>
<td>Management</td>
</tr>
<tr>
<td>person to drive with?</td>
<td></td>
</tr>
<tr>
<td>Some drivers need to pay more attention to driving.</td>
<td>Acknowledgement</td>
</tr>
<tr>
<td>How do you think that you are driving?</td>
<td>Management</td>
</tr>
</tbody>
</table>

**Part two had 3 Acknowledgement prompts, 4 Conversational prompts, and 1 Suggestion prompt**

<table>
<thead>
<tr>
<th>Prompt</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>Parts of this road are under construction.</td>
<td>Acknowledgement</td>
</tr>
<tr>
<td>I like driving on mountain roads, what is your favourite road to</td>
<td>Management</td>
</tr>
<tr>
<td>drive?</td>
<td></td>
</tr>
<tr>
<td>There is a traffic jam ahead; if you turn left you might avoid it.</td>
<td>Suggestion</td>
</tr>
<tr>
<td>How are you feeling right now</td>
<td>Management</td>
</tr>
</tbody>
</table>
This road has many construction zones. Acknowledgement
What kinds of things do you think about while driving? Management
Some drivers should pay more attention to driving. Acknowledgement
Do find it stressful talking to people while driving? Management

Part three had 3 Acknowledgement prompts, 4 Conversational prompts, and 1 Suggestion prompt

<table>
<thead>
<tr>
<th>Prompt</th>
<th>Response</th>
</tr>
</thead>
<tbody>
<tr>
<td>I would love to ride in one of the new Mini’s, what is your favorite</td>
<td>Management</td>
</tr>
<tr>
<td>car to drive?</td>
<td></td>
</tr>
<tr>
<td>There is an accident ahead; if you turn left here you might avoid it.</td>
<td>Suggestion</td>
</tr>
<tr>
<td>What do you think about this car?</td>
<td>Management</td>
</tr>
<tr>
<td>This stretch of road is often slippery.</td>
<td>Acknowledgement</td>
</tr>
<tr>
<td>This road is too narrow and windy for slow vehicles.</td>
<td>Acknowledgement</td>
</tr>
<tr>
<td>I find that driving on narrow road makes me stressed what kind of</td>
<td>Management</td>
</tr>
<tr>
<td>roads make you feel stressed?</td>
<td></td>
</tr>
<tr>
<td>This stretch of road often has a problem with fog</td>
<td>Acknowledgement</td>
</tr>
<tr>
<td>That sure was miserable, what’s your strategy for coping with rain</td>
<td>Management</td>
</tr>
<tr>
<td>and fog while driving?</td>
<td></td>
</tr>
</tbody>
</table>

Part four had 3 Acknowledgement prompts, 4 Conversational prompts, and 1 Suggestion prompt

7.1.2 Speech Prompts for Happy – Sad Drivers

After the initial prompt, the prompts listed in the table below were placed at random locations in the driving course. All drivers experienced the prompts in the exact same order.

Hi, My name is Chris and I will be your virtual passenger for today. We are going to be driving on a coastal road, one that I have travelled many times before, but one that might be new to you. The trips should not take long, 10 or 15 minutes, let’s get going.

<table>
<thead>
<tr>
<th>Prompt</th>
<th>Response</th>
</tr>
</thead>
<tbody>
<tr>
<td>How are you doing</td>
<td>How many vehicles do you think you have passed in your lane</td>
</tr>
<tr>
<td>How do you think that the car is performing</td>
<td>Do you remember what the last posted speed limit was</td>
</tr>
<tr>
<td>Where do you drive most frequently in</td>
<td>For how long do you think you have been</td>
</tr>
<tr>
<td>Question</td>
<td>Answer</td>
</tr>
<tr>
<td>-------------------------------------------------------------------------</td>
<td>------------------------------------------------------------------------</td>
</tr>
<tr>
<td>your car</td>
<td>driving today</td>
</tr>
<tr>
<td>Do you generally like to drive at, below or above the speed limit</td>
<td>I saw the new Matrix a couple of days ago and it was OK, what was the last good movie that you have seen</td>
</tr>
<tr>
<td>There seem to be a lot of bridges, overpasses and tunnels around here,</td>
<td>Do you remember if we have passed the sign for Lucia</td>
</tr>
<tr>
<td>how many tunnels do you think we have gone through so far.</td>
<td></td>
</tr>
<tr>
<td>Have we come to the turnoff for Morro Bay yet</td>
<td>Approximately how many oncoming cars have passed you so far</td>
</tr>
<tr>
<td>I get stressed in traffic almost every day, how often are you stressed</td>
<td>What do you think about the car</td>
</tr>
<tr>
<td>by traffic problems</td>
<td></td>
</tr>
<tr>
<td>How do you feel about the car</td>
<td>Can you tell me what you like about the car</td>
</tr>
<tr>
<td>What do you dislike about the car</td>
<td>Don’t you think that these lanes are a little too narrow</td>
</tr>
<tr>
<td>What do you think about the driving conditions</td>
<td>How do you think you are driving</td>
</tr>
<tr>
<td>I usually study guidebooks when I have free time, what do you like to</td>
<td>Other than a speed limit posting, describe the last road sign you saw</td>
</tr>
<tr>
<td>do in your spare time</td>
<td></td>
</tr>
<tr>
<td>Can you describe the last car you saw from oncoming traffic</td>
<td>One of my favourite parts of this drive is the light house, be sure to tell me what you think of it once you see it</td>
</tr>
<tr>
<td>How are you feeling right now</td>
<td>How do you usually feel when you are driving</td>
</tr>
<tr>
<td>Have we passed a couple of white houses on the right side? The first</td>
<td>If you could change something about your car, what would you change</td>
</tr>
<tr>
<td>time I drove here I had to stop at the houses and ask for directions.</td>
<td></td>
</tr>
<tr>
<td>When was the last time you got lost driving somewhere</td>
<td></td>
</tr>
<tr>
<td>If you could change something about your driving what would you change</td>
<td>I like to drive with people that talk to me, can you tell me about your favourite person to drive with</td>
</tr>
<tr>
<td>I don’t enjoy driving with people that don’t talk to me, who is your</td>
<td>What kinds of things do you think about when you are driving</td>
</tr>
<tr>
<td>least favourite person to drive with</td>
<td></td>
</tr>
</tbody>
</table>
One of my favourite road trips is to driving through the southwest to the grand canyon, can you tell me about your favourite road trip destination?
I listen to Madonna’s new CD a lot, can you tell me about your favourite CD?

What do you think about the weather we have been having?
It sure is sunny today, what is your strategy for coping with bright conditions while driving?

7.1.3 Questionnaire - Prior Driving Experience

All participants in both studies answered the questionnaire below before the driving session. This questionnaire was designed by the team of faculty and students working with driving simulator studies at Stanford University. Most questions are general, but others such as driving experience and video game experience were explicitly targeted for driving studies.

What is your gender?
What is your age?
On average how many hours of computer games/video games do you play per day?
Which specific computer games/video games are you most familiar with?
How many years have you been driving?
How many accidents have you been in?
How many speeding tickets have you received?
Have you used a cell phone/mobile phone while driving?
Do you normally drive with passengers in the car?
Do you sometimes get upset while driving?
Do you communicate your anger with other drivers while driving?
Do you like driving?
Do you drive in rush hour?
Do you take long road trips?
How many miles/kilometers do you normally drive per week?
Where do you normally drive? Freeway/Motorway, urban or city?

7.1.4 Questionnaire – 32 Term DES

This questionnaire, Differential Emotional Scale – DES (Izard 1977), was used in both studies in this chapter. Participants were asked how well the affective adjectives matched their current feelings using a 10 point Likert scale, where 1 equals describes very poorly and 10 equals describes very well.
Appendices

<table>
<thead>
<tr>
<th>Amused</th>
<th>Alert</th>
</tr>
</thead>
<tbody>
<tr>
<td>Guilty</td>
<td>Hostile</td>
</tr>
<tr>
<td>Frightened</td>
<td>Happy</td>
</tr>
<tr>
<td>Upset</td>
<td>Fearful</td>
</tr>
<tr>
<td>Astonished</td>
<td>Ashamed</td>
</tr>
<tr>
<td>Unhappy</td>
<td>Distressed</td>
</tr>
<tr>
<td>Dislike</td>
<td>Attentive</td>
</tr>
<tr>
<td>Interested</td>
<td>Distaste</td>
</tr>
<tr>
<td>Afraid</td>
<td>Aggressive</td>
</tr>
<tr>
<td>Timid</td>
<td>Disgusted</td>
</tr>
<tr>
<td>Shocked</td>
<td>Amazed</td>
</tr>
<tr>
<td>Delighted</td>
<td>Enthusiastic</td>
</tr>
<tr>
<td>Scared</td>
<td>Surprised</td>
</tr>
<tr>
<td>Sad</td>
<td>Curious</td>
</tr>
<tr>
<td>Repulsion</td>
<td>Embarrassed</td>
</tr>
<tr>
<td>Mad</td>
<td>Anger</td>
</tr>
</tbody>
</table>

7.1.5 Questionnaire – Driving Experience

The questionnaires below were used for both studies on emotional drivers. The questionnaires were designed by faculty and students working with driving simulator studies at Stanford University. In these questionnaires participants were asked how well the adjectives/statements matched their experience using a 10-point Likert scale, where 1 equals describes very poorly and 10 equals describes very well.

How well do the following words describe how you felt while driving with the in-vehicle system?

<table>
<thead>
<tr>
<th>Safe</th>
<th>Calm</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tired</td>
<td>Tense</td>
</tr>
<tr>
<td>Active</td>
<td>Strained</td>
</tr>
<tr>
<td>At ease</td>
<td>Frustrated</td>
</tr>
<tr>
<td>Comfortable</td>
<td>Passive</td>
</tr>
<tr>
<td>Drowsy</td>
<td>Confident</td>
</tr>
<tr>
<td>Nervous</td>
<td>Indecisive</td>
</tr>
<tr>
<td>Relaxed</td>
<td>Secure</td>
</tr>
<tr>
<td>Fatigued</td>
<td>Steady</td>
</tr>
<tr>
<td>Confused</td>
<td></td>
</tr>
</tbody>
</table>
How well do the following statements describe the In-Vehicle systems influence on your driving?

- I was a more alert driver
- I was a more careful driver
- I was a more confident driver
- I was a more distracted driver
- I was a safer driver
- I was a more aggressive driver

The questionnaires that asked participants about the perceived quality of the in-vehicle system and the car contained the exact same statements.

How well do the following words describe how you feel about the in-vehicle system? or How well do the following words describe how you feel about the car?

- Trustworthy
- Annoying
- Intelligent
- I would buy
- Well-designed
- High in quality
- Helpful
- Authoritative
- I would use again
- I would recommend to a friend or family member
- Friendly
- Boring
- Frustrating
- Fun to use
- I want to have
- Confusing
- Condescending
- likable
- reliable

7.1.6 Questionnaire – Impact of horn honks on driving

This questionnaire was used for the second study, and it was designed by faculty and students at Stanford University. Participants were asked how well the statements matched their experience on a 10-point Likert scale, where 1 equals describes very poorly and 10 equals describes very well.

- Responding to the horn helped my driving performance
- I felt I did a good job driving
- I felt I was a safe driver
- I thought it was difficult to respond to the horn and remain focused
- I responded to the horn promptly after the cues
- It was easy to respond to the horn
- Responding to the horn hurt my driving performance
### 7.2 Appendix B

#### 7.2.1 Prompts for Accurate and Inaccurate Information in cars

The following tables contain the prompts used for the study on accuracy of information. Each table contains prompts used for a specific level of accuracy. Please note that the exact same prompts are used in each table even though the order might be different.

**Prompts for 100% accuracy**

<table>
<thead>
<tr>
<th>Prompt</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>cP1</td>
<td>Make sure to obey the speed limit. The last posted speed was 35 miles per hour.</td>
</tr>
<tr>
<td>cP2</td>
<td>You are approaching an intersection.</td>
</tr>
<tr>
<td>cP3</td>
<td>Caution, watch for merging traffic.</td>
</tr>
<tr>
<td>cP4</td>
<td>You are approaching an intersection.</td>
</tr>
<tr>
<td>cP5</td>
<td>You can increase your speed. The current speed limit is 65 miles per hour.</td>
</tr>
<tr>
<td>cP6</td>
<td>Caution, there is heavy fog ahead.</td>
</tr>
</tbody>
</table>

**Prompts for Om1 accuracy**

<table>
<thead>
<tr>
<th>Prompt</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>cP7</td>
<td>You are approaching an intersection.</td>
</tr>
<tr>
<td>Cp8</td>
<td>Current speed limit is 45 miles per hour.</td>
</tr>
<tr>
<td>Cp30</td>
<td>Warning, watch for oncoming traffic in your lane.</td>
</tr>
<tr>
<td>Cp9</td>
<td>Caution, the road is slippery.</td>
</tr>
<tr>
<td>Cp10</td>
<td>Warning, there is a car stopped in your lane.</td>
</tr>
</tbody>
</table>

**Prompts for Om2 accuracy**

<table>
<thead>
<tr>
<th>Prompt</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cp11</td>
<td>Warning, there is a fallen tree in the middle of the road.</td>
</tr>
<tr>
<td>Cp12</td>
<td>Warning, the vehicle ahead of you has stopped.</td>
</tr>
<tr>
<td>Cp13</td>
<td>Caution, there are slow-moving vehicles ahead.</td>
</tr>
<tr>
<td>Cp14</td>
<td>The road widens ahead.</td>
</tr>
<tr>
<td>Cp15</td>
<td>Caution, you are entering a high-wind area.</td>
</tr>
</tbody>
</table>

**Prompts for Om3 accuracy**

<table>
<thead>
<tr>
<th>Prompt</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cp16</td>
<td>Slow for the road construction ahead.</td>
</tr>
<tr>
<td>Cp17</td>
<td>Current speed limit is 60 miles per hour.</td>
</tr>
<tr>
<td>Cp18</td>
<td>You are approaching a traffic light.</td>
</tr>
<tr>
<td>Cp19</td>
<td>Current speed limit is now 45 miles per hour, adjust your speed accordingly.</td>
</tr>
<tr>
<td>Cp20</td>
<td>You are approaching an intersection.</td>
</tr>
<tr>
<td>Cp21</td>
<td>Slow for the road construction ahead.</td>
</tr>
</tbody>
</table>

**Prompts for Om4 accuracy**

<table>
<thead>
<tr>
<th>Prompt</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cp22</td>
<td>Caution, there are several intersections ahead.</td>
</tr>
<tr>
<td>Cp23</td>
<td>You are approaching an intersection.</td>
</tr>
</tbody>
</table>

**Prompts for Om5 accuracy**

<table>
<thead>
<tr>
<th>Prompt</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cp24</td>
<td>Caution, you are entering a school zone. The speed limit is 25 miles per hour.</td>
</tr>
<tr>
<td>Cp25</td>
<td>Current speed limit is 55 miles per hour.</td>
</tr>
</tbody>
</table>
Appendices

Cp31  Caution, emergency vehicles approaching, please move to the right side of the road
Cp26  Make sure that you are going an appropriate speed. The speed limit is 45 miles
Om6  Cp27  There is a stop sign ahead.
Cp28  Current speed limit is 55 miles per hour.
Cp29  You are approaching an intersection and a gas station.
Cp33  Caution, watch for merging traffic.
Om13  Cp32  Warning, the vehicle ahead of you has stopped. Proceed with caution.

Prompts for 88% accuracy (90%)
Cp1    Make sure to obey the speed limit. The last posted speed was 35 miles per hour.
Cp2    You are approaching an intersection.
Cp3    Caution, watch for merging traffic.
Cp4    You are approaching an intersection.
Cp5    You can increase your speed. The current speed limit is 65 miles per hour.
Cp6    Caution, there is heavy fog ahead.
Cp7    You are approaching an intersection.
Cp8    Current speed limit is 45 miles per hour.
Cp30   Warning, watch for oncoming traffic in your lane.
Cp9    Caution, the road is slippery.
Cp10   Warning, there is a car stopped in your lane.
Omitted  Cp11

Inserted  Cp24  Caution, you are entering a school zone. The speed limit is 25 miles per hour.
Cp12   Warning, the vehicle ahead of you has stopped.
Cp13   Caution, there are slow-moving vehicles ahead.
Cp14   The road widens ahead.
Cp15   Caution, you are entering a high-wind area.
Cp16   Slow for the road construction ahead.
Cp17   Current speed limit is 60 miles per hour.
Cp18   You are approaching a traffic light.
Cp19   Current speed limit is now 45 miles per hour, adjust your speed accordingly.
Cp20   You are approaching an intersection.
Cp21   Slow for the road construction ahead.
Cp22   Caution, there are several intersections ahead.
Inserted  
Cp11  Warning, there is a fallen tree in the middle of the road.
Cp23  You are approaching an intersection

Omitted  
Cp24
Cp25  Current speed limit is 55 miles per hour.
CP31  Caution, emergency vehicles approaching, please move to the right side of the road
Cp26  Make sure that you are going an appropriate speed. The speed limit is 45 miles
Cp27  There is a stop sign ahead.
Cp28  Current speed limit is 55 miles per hour.
Cp29  You are approaching an intersection and a gas station.
Cp33  Caution, watch for merging traffic.
Cp32  Warning, the vehicle ahead of you has stopped. Proceed with caution.

Prompts for 76% accuracy (80%)
Cp1  Make sure to obey the speed limit. The last posted speed was 35 miles per hour.
Cp2  You are approaching an intersection.
Cp3  Caution, watch for merging traffic.
Cp4  You are approaching an intersection.
Cp5  You can increase your speed. The current speed limit is 65 miles per hour.
Cp6  Caution, there is heavy fog ahead.

Omitted  
Cp7
Cp8  Current speed limit is 45 miles per hour.
Cp30  Warning, watch for oncoming traffic in your lane
Cp9  Caution, the road is slippery.
Cp10  Warning, there is a car stopped in your lane

Omitted  
Cp11
Inserted  
Cp7  You are approaching an intersection.
Cp12  Warning, the vehicle ahead of you has stopped.
Cp13  Caution, there are slow-moving vehicles ahead.

Inserted  
Cp16  Slow for the road construction ahead.
Cp14  The road widens ahead.
Cp15  Caution, you are entering a high-wind area.

Omitted  
Cp16
Inserted  
Cp24  Caution, you are entering a school zone. The speed limit is 25 miles per hour.
Cp17  Current speed limit is 60 miles per hour.
Cp18  You are approaching a traffic light.
Cp19  Current speed limit is now 45 miles per hour, adjust your speed accordingly.
Cp20  You are approaching an intersection.
Cp21  Slow for the road construction ahead.
Cp22  Caution, there are several intersections ahead.

Inserted  Cp11  Warning, there is a fallen tree in the middle of the road.
Cp23  You are approaching an intersection

Omitted  Cp24
Cp25  Current speed limit is 55 miles per hour.
CP31  Caution, emergency vehicles approaching, please move to the right side of the road
Cp26  Make sure that you are going an appropriate speed. The speed limit is 45 miles
Cp27  There is a stop sign ahead.
Cp28  Current speed limit is 55 miles per hour.
Cp29  You are approaching an intersection and a gas station.
Cp33  Caution, watch for merging traffic.
Cp32  Warning, the vehicle ahead of you has stopped. Proceed with caution.

Prompts for 64% accuracy (70%)
Cp1  Make sure to obey the speed. The last posted speed was 35 miles per hour.
Cp2  You are approaching an intersection.
Cp3  Caution, watch for merging traffic.
Cp4  You are approaching an intersection.
Cp5  You can increase your speed. The current speed limit is 65 miles per hour
Cp6  Caution, there is heavy fog ahead.

Omitted  Cp7
Cp8  Current speed limit is 45 miles per hour.
Cp30  Warning, watch for oncoming traffic in your lane
Cp9  Caution, the road is slippery.
Cp10  Warning, there is a car stopped in your lane

Omitted  Cp11
Inserted  Cp7  You are approaching an intersection.
Cp12  Warning, the vehicle ahead of you has stopped.
Cp13  Caution, there are slow-moving vehicles ahead.
Social and Emotional Characteristics of Speech-based In-Vehicle Information Systems: Impact on Attitude and Driving Behaviour

Inserted  
 Cp16  Slow for the road construction ahead.
 Cp14  The road widens ahead.
 Inserted  
 Cp27  There is a stop sign ahead.
 Cp15  Caution, you are entering a high-wind area.
 Omitted  
 Cp16
 Inserted  
 Cp24  Caution, you are entering a school zone. The speed limit is 25 miles per hour.
 Cp17  Current speed limit is 60 miles per hour.
 Cp18  You are approaching a traffic light.
 Cp19  Current speed limit is now 45 miles per hour, adjust your speed accordingly.
 Cp20  You are approaching an intersection.
 Cp21  Slow for the road construction ahead.
 Omitted  
 Cp22
 Inserted  
 Cp11  Warning, there is a fallen tree in the middle of the road.
 Cp23  You are approaching an intersection
 Omitted  
 Cp24
 Cp25  Current speed limit is 55 miles per hour.
 CP31  Caution, emergency vehicles approaching, please move to the right side of the road
 Cp26  Make sure that you are going an appropriate speed. The speed limit is 45 miles
 Omitted  
 Cp27
 Cp28  Current speed limit is 55 miles per hour.
 Cp29  You are approaching an intersection and a gas station.
 Cp33  Caution, watch for merging traffic.
 Cp32  Warning, the vehicle ahead of you has stopped. Proceed with caution.
 Inserted  
 Cp22  Caution, there are several intersections ahead.

7.2.2 Questionnaire - Prior Driving Experience

All participants answered the following questionnaire before the driving sessions. This questionnaire was designed by the team of faculty and students working with driving simulator studies at Stanford University.

What is your gender?
What is your age?
On average how many hours of computer games/video games do you play per day
Which specific computer games/video games are you most familiar with
How many years have you been driving?
How many accidents have you been in?
How many speeding tickets have you received?
Have you used a cell phone/mobile phone while driving?
Do you normally drive with passengers in the car?
Do you sometimes get upset while driving?
Do you communicate your anger with other drivers while driving?
Do you like driving?
Do you drive in rush hour?
Do you take long road trips?
How many miles/kilometers do you normally drive per week?
Where do you normally drive? Freeway/Motorway, urban or city

7.2.3 Questionnaire – Impact of and Quality of In-Vehicle system and Car

All participants answered the following questionnaires. This questionnaire was designed by the team of faculty and students working with driving simulator studies at Stanford University. Please note the same statements were used to assess quality of in-vehicle system and car. All questionnaires below were designed with 10 point Likert scales, where “describes very poorly” equals 1 and “describes very well” equals 10.

How well do the following words describe how you feel about the voice system or
How well do the following words describe how you feel about the car respectively?

Fun to use Poorly designed
High in quality I would discourage family and friends from using it
Inferior quality I would use it again
I would buy it Boring to use
I would not pay for it I want to have it
Well designed I have no desire for it
I would recommend it to a friend or family member I would not use it

7.2.4 Questionnaire – Credibility and Trust

These questionnaires are copyrighted material; please see the Individualized Trust Scale and scales for Source Credibility in Communication Research Measures: A

### 7.3 Appendix C

#### 7.3.1 Speech Prompts for Study matching age of voice with age of driver

There were 25 speech prompts used by the in-car information system, 10 were informational, 10 were warnings, and 5 were suggestions.

**Informational Prompts:**
1. The current speed limit is 60 miles an hour
2. You are approaching an intersection
3. Road ahead narrows and starts to wind
4. You are approaching an intersection
5. You are entering a school zone.
6. The current speed limit is 60 miles an hour
7. Road narrows
8. Stop sign ahead.
9. The current speed limit is 30 miles per hour
10. You are entering a crosswinds area

**Warning Prompts:**
1. Caution, there is thick fog ahead
2. Warning, there is a fallen tree in the road ahead
4. Warning, road works ahead
5. Beware the vehicle ahead of you appears to have broken down.
6. Caution, the road is very slippery.
9. Caution, there is overtaking traffic in your lane
10. Caution, Police is approaching from behind

**Suggestions:**
1. The police use radar here, you might need to slow down
2. There is heavy traffic ahead; if you turn left you can avoid it.
3. There is an accident ahead, if you turn right you might avoid it.
4. There is traffic congestion ahead; if you turn left you can avoid it.
5. There is traffic congestion ahead, if you turn right you might avoid it.
7.3.2 Questionnaire - Prior Driving Experience

All participants, both young drivers and older adult drivers answered the following questions before the driving session. This questionnaire was designed by the team of faculty and students working with driving simulator studies at Stanford University.

What is your gender?
What is your age?
On average how many hours of computer games/video games do you play per day
Which specific computer games/video games are you most familiar with
How many years have you been driving?
How many accidents have you been in?
How many speeding tickets have you received?
Have you used a cell phone/mobile phone while driving?
Do you normally drive with passengers in the car?
Do you sometimes get upset while driving?
Do you communicate your anger with other drivers while driving?
Do you like driving?
Do you drive in rush hour?
Do you take long road trips?
How many miles/kilometers do you normally drive per week?
Where do you normally drive? Freeway/Motorway, urban or city

The following questions were added to the questionnaire for the older adults – based on AAA 55+ (American-Automobile-Association-Foundation-for-Traffic-Safety 1994). Older adult participants were asked to answer the questions on a 3 point scale ranging from Always to Sometimes to Never.

I signal and check to the rear when I change lanes
I wear a seat belt
I try to stay informed on changes about driving and highway regulations
Intersections bother me because there is so much to watch from all directions
I find it difficult to decide when to join traffic on a busy main road
I think that I am slower than I used to be in reacting to dangerous driving situations
When I am really upset, I show it in my driving
My thoughts wander when I am driving
Traffic situations make me angry
I get regular eye checks to keep my vision at its sharpest
I check with my doctor or pharmacist about effects of my medications on driving ability (If you do not take any medication skip this question)
I try to stay abreast of current information on health practices and habits
My children, other family members or friends are concerned about my driving ability
I avoid driving when it is dark
Social and Emotional Characteristics of Speech-based In-Vehicle Information Systems: Impact on Attitude and Driving Behaviour

I find it easy to make left turns
I find it easy to make right turns
I am comfortable driving at high speeds on main roads

7.3.3 Questionnaire - Quality of Voice

Participants listening to the two voices were asked to assess the quality of the voice based on the following set of contrasting adjectives based on a 10 point Likert scale. The questionnaire was designed by me after consulting with Christian Jones. I failed to find a standard questionnaire for quality of voice. Questionnaires in this area seem more targeted towards quality of service than quality of voice per se.

Intelligible  Inarticulate
Whispering-quiet  Shouting-Loud
Harmonious  Inharmonious
Hard  Soft
Rich in tones  Flat
Spiritless  Vibrant
High Pitch  Low Pitch
Slow  Fast
Hoarse  Clear
Sweet-Warm  Dry-Harsh
Breathy  Non-breathy
Hesitant  Fluent
Calm  Stressed
Slurred  Enunciated

7.3.4 Questionnaire – 20-term DES

In the first study, participants were asked to assess the emotional colouring the voice using the questionnaire below. In the second study, participants were asked before and after the driving session how well the affective adjectives matched their current feelings using a 10 point Likert scale, where 1 equals “describes very poorly” and 10 equals “describes very well”.

Angry  Upset
Hostile  Sad
Aggressive  Fearful
Mad  Scared
Enthusiastic  Anxious
Happy  Afraid
Excited  Calm
7.3.5 Questionnaire – Trust and Credibility of Voice and In-vehicle system and Homophily

These questionnaires were used in both studies and are copyrighted material. Please see the Individualized Trust Scale and scales for Source Credibility in Communication Research Measures: A Sourcebook, by Rubin, R. B., P. Palmgreen, and H. E. Sypher (Rubin, Palmgreen et al. 1994).

7.3.6 Questionnaire – Quality of the In-Vehicle System and car

The following questionnaires were used in the second study after the driving session. This questionnaire was designed by the team of faculty and students conducting driving simulator studies and Stanford University. Please note that the exact same statements were used to assess both the quality of the in-vehicle system and the quality of the car.

How well do the following words describe how you feel about the voice system or How well do the following words describe how you feel about the car respectively?

| Fun to use | Poorly designed |
| High in quality | I would discourage family and friends from using it |
| I would not buy it or pay for it | I would use it again |
| I want to have it | |

7.3.7 Questionnaire – influence of in-vehicle system and AAA 55+

All participants were asked to rate the influence of the system based on the following statements on a 10-point Likert scale with 1 = describes very poorly and 10 = describes very well. The first questionnaire was designed by faculty and students working with driving simulator studies at Stanford University.

An aggressive driver
Enjoy the driving experience
Feel Confused
Drive carefully
An inattentive driver
In addition to the above, older adult participants were also asked to rate how the in-vehicle system influenced their driving on a 10-point Likert scale, with 1 = describes very poorly and 10 = describes very well. This part of the questionnaire was designed based on the AAA 55+ pamphlet (American-Automobile-Association-Foundation-for-Traffic-Safety 1994).

- Watch more carefully at intersections
- React faster to dangerous driving situations
- More comfortable driving at faster speeds
- A better driver in low-visibility condition