Master Thesis in Interaction Design

A Design Review of a Voice Control Prototype for Volvo's Navigation System

Annika Harup & Emelie Normand

Gothenburg, Sweden 2004

IT University of Göteborg

Chalmers | Göteborgs Universitet

Chalmers Department of Computing Science

REPORT NO. 2004:45

A Design Review of a Voice Control Prototype for Volvo's Navigation System

ANNIKA HARUP & EMELIE NORMAND

Master program "Human Computer Interaction and Interaction Design"
Department of Computer Science at Chalmers University of Technology
IT UNIVERSITY OF GÖTEBORG
GÖTEBORG UNIVERSITY AND CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden 2004

”A Design Review of a Voice Control Prototype for Volvo’s Navigation System”
ANNIKA HARUP & EMELIE NORMAND

Report no 2004:45
ISSN: 1651-4769
Department of Computer Science
IT University of Göteborg
Göteborg University and Chalmers University of Technology
P O Box 8718
SE – 402 75 Göteborg
Sweden
Telephone + 46 (0)31-772 4895

Tryckeriet Matematiskt centrum
Göteborg, Sweden 2004

”A Design Review of a Voice Control Prototype for Volvo’s Navigation System”

ANNIKA HARUP & EMELIE NORMAND

Department of Computer Science
IT University of Gothenburg
Gothenburg University and Chalmers University of Technology

ABSTRACT

As technology evolves new opportunities arise to create innovating applications that can aid in every-day life. In the car industry it has become more common to apply new technology to meet the needs that a driver might have in an *in-car* environment. An example of such an application is an electronic navigation system that can guide the driver to a preferred destination or supply the driver with useful traffic information. However, these navigation systems can be very complex and difficult to interact with as the driver needs to keep his attention on the road. One way of letting the driver keep his hands on the wheel and eyes on the road is to use voice control to interact with the navigation system. The purpose of this thesis is to evaluate a voice control prototype that was created for Volvo Car’s navigation system.

In this thesis the voice control prototype has been evaluated with a user-centred design method in mind. A usability test has been conducted to study the interaction between the user and the system. We have questioned the use of only one modality when interacting with the system or if using a multimodal way of interacting is to prefer.

The results of our evaluation shows that it is important that the Speech User Interface is adapted to the user’s current situation, in this case an *in-car* environment where the driver’s mental capacity at times can be very limited. To avoid that the system constitutes a traffic hazard the menu structure, to give one example, should be designed so that it contributes with only a minimum amount of mental workload. The results of our evaluation also showed that the test participants sometimes tried to use more than one modality when solving a task. This can indicate that some of the tasks are more easily solved manually, some with voice control and some tasks are more easily solved with a combination of both modalities.

Keywords: Speech Technology, Speech Recognition, Speech Synthesis, Voice Control, Speech User Interface, User Centred Design, User Test, Navigation Systems

”A Design Review of a Voice Control Prototype for Volvo’s Navigation System”

ANNIKA HARUP & EMELIE NORMAND

Institutionen för Datavetenskap
IT Universitetet i Göteborg
Göteborg Universitet and Chalmers Tekniska Högskola

SAMMANFATTNING

I takt med teknikens utveckling uppstår det nya möjligheter att skapa applikationer som kan underlätta vardagen. Inom bilindustrin blir det allt vanligare att tillämpa ny teknik för att tillgodose de behov föraren kan ha i bilmiljö. Ett exempel på en sådan tillämpning är elektroniska navigationssystem som hjälper föraren att hitta en önskad plats eller tillhandahålla föraren med väsentlig trafikinformation. Dessa kan dock vara komplexa och svåra att interagera med då föraren behöver vara uppmärksam på trafiken. Ett sätt att frigöra förarens händer och syn till köruppgiften är att använda sig av röststyrning för att interagera med navigationssytemet. Syftet med denna rapport är att utvärdera en röststyrningsprototyp som gjorts för Volvo Car’s navigationssystem.

I den här rapporten har röststyrningsprototypen utvärderats utifrån en användarcentrerad designmetod. Ett användartest har gjorts för att studera interaktionen mellan användare och system. Vi har ställt oss frågande till om man bara bör använda sig av en modalitet när man interagerar med systemet eller om ett multimodalt interaktionssätt är att föredra.

Resultatet av vår undersökning visar att det är viktigt att talgränssnittet är anpassat till användarens situation, i detta fall en bilmiljö där förarens mentala kapacitet stundtals kan vara väldigt begränsad. För att systemet inte ska utgöra en trafiksäkerhetsrisk måste exempelvis menystrukturen vara designad så att den medför minimal mental arbetsbelastning. Det visade sig även att testdeltagarna gjorde försök att interagera med systemet genom flera modaliteter då de löste en testuppgift. Detta kan indikera att vissa uppgifter löses enklast manuellt, vissa med röst och andra med en kombination av bägge.

ACKNOWLEDGMENT

We would like to thank our supervisors; Patricija Jaksetic at the IT-university for all the help, support and guidance given to us in order to write this thesis, and Johannes Agardh at Volvo Car Corporation for answering all our questions, supplying us with valuable information and for giving us all that practical help.

We would also like to thank the staff within the department Comfort and Driver Information at Volvo Car Corporation who have helped us, made us feel welcome and put up with us during our stay at the department and especially thanks to Camilla Håkansson for taking all that time with us.

To all the staff that participated in our user test: thank you all for your time and admirable patience during our tests! Your contribution was very important for our thesis!

We would like to thank Lewis Mowatt and James Aggiss at Jaguar for their help and for lending us their Jaguar AVP prototype analysed in this thesis.

Last but not least we would like to thank our family and friends for all support and patience.

Annika Harup Emelie Normand

Gothenburg, May 2004

TABLE OF CONTENTS

1 INTRODUCTION

This Master Thesis is the final work at the M. Sc. Program in HCI/Interaction design at the IT-University, a part of Chalmers University of Technology. The work includes 20 credits which is equivalent to 20 weeks of work. The thesis is carried out at Volvo Car Corporation, at the unit for Driver Information & Interaction design.

Volvo has been developing a prototype for voice control for the infotainment system (audio, phone, navigation and mobile office). The aim with the thesis is to make an improvement plan for Volvo’s voice control of the navigation system (there is only a voice control prototype for the navigation system). To investigate how the system can be improved a user test on the voice control was conducted. The test was conducted in an in-vehicle environment on the current voice control prototype. An analysis of another voice control prototype for operating a navigation system has also been made. This analysis was based on a cognitive walkthrough of the prototype.

At start a decision was to be taken at Volvo about how to proceed with the Speech User Interface (SUI) (see Lexicon) for the navigation system. There was a choice between two different systems and there are some major differences between them. The idea was to use the results of this report as input.

A major advantage with voice control for in-vehicle information systems is that using voice control enables the driver to keep his hands on the steering wheel and the eyes on the road to a greater extent. This could be a way to enhance driving safety on the condition that the communication between the driver and the system works well. The speech interface must provide the driver with an easy way to communicate his message to the information system. The interface also has to provide the driver with comprehensible output.

The goal of communication is to achieve a mutual understanding of the message. This goal has been reached when the sender and the receiver makes the same interpretation of the message (Weinschenk, 2000*).* (See Communication in the Lexicon)

Voice control uses the speech elements within verbal communication to give input to the navigation system. Verbal communication also includes written language which is used to give output from the navigation system in addition to speech. In this thesis focus lies on spoken communication; that is dialogue. A dialogue is a kind of interaction or a process where two or more parties continuously are reacting on each other’s actions. It is important that in a dialogue both sides show that they are listening and understanding, to do this, people use *turn-taking* and *feedback.*

Since communication is a way of collaboration there has to be rules to form a dialogue; this is called turn taking. These rules make the speakers know when to speak and when to listen. Turn-taking can be compared to a negotiation process where the negotiators decide who to speak next (Baber, Noyes, 1993). The person currently speaking can for example indicate when he is ready to let the other person continue the dialogue by lowering his voice pitch and looking into the listener’s eyes. On the other hand if the listener wishes to add something to the dialogue while the other person is still talking he can try to interrupt the speaker verbally. Gestures like frequently nodding are another way to indicate that a person wishes to enter the dialogue. A third way is to look intensely into the speaker’s eyes (something that can be found uncomfortable for both parts (Allwood, 1986). When turn taking processes are inhibited the communication is affected. This may be manifested in longer pauses when switching between speakers or in communication structure break down (McKinlay, 1993). The design challenge is to develop a system that can accurately detect when the speaker has finished and quickly move on to the next step in the dialogue (Gardner-Bonneau, 1999).

Dialogue feedback can display itself in different ways: verbal, a non verbal sound, a gesture and/or other type of body language. It can be used by the receiver to reassure the sender that the receiver is listening. By using feedback the receiver can indicate if he has understood the message or not and if he agrees with it. Without feedback the communication process breaks down (Weinschenk, 2000).

When designing a Speech User Interface it is essential to take in consideration the importance of a smooth turn-taking process and how to use system feedback to support the turn-taking between human and machine so breakdowns are avoided.

In a human to human dialogue 75-90% of the message is conveyed with nonverbal signals (Ahrenfelt, Berner, 1999). Nonverbal information does not contain words but can be a great variety of signals, some are clothes and touch, but perhaps the most important nonverbal signals are intonation, facial expressions, gestures and other types of body language.

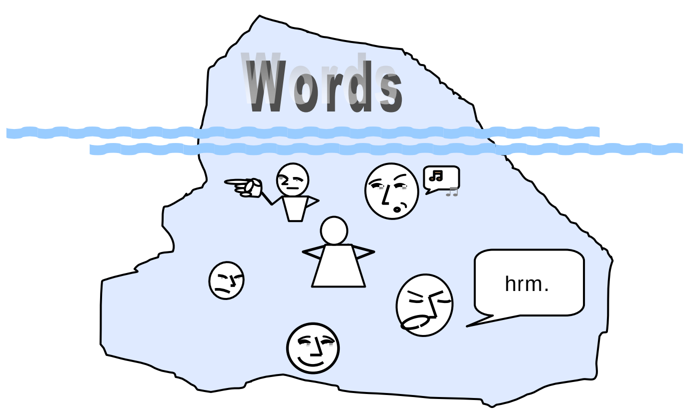


Figure 1: Shows the message as an isberg were the message is more than mearly words.

Without intonation, that is adjustments in pitch, the spoken language sounds monotone and a lot of information is lost. Speakers use for example intonation to indicate questions, to make statements and to emphasize something.

As already mentioned gestures are often used simultaneously as spoken language to enable the turn-taking process and to give feedback. Deictic gestures like pointing can often precede speech, for example if someone points to a car and says "*that one is mine*".

Speech User Interfaces can not interpret non verbal communication signals like those described above; therefore the users have to be aware of this when talking to a Speech User Interface and communicate their message through speech only (and perhaps by manual devices). On the other hand non verbal signals, such as intonation, can be used to make it easier for the user to understand speech output produced by the Speech User Interface.

1.1 VOLVO CAR CORPORATION

When Assar Gabrielsson and Gustaf Larsson, the founders of Volvo, laid down the broad outlines for the first Volvo car, the user was put in focus. Volvo began manufacturing cars in 1927 and today, more than 75 years later, both society and cars look different but Volvo's focus when manufacturing cars is still the same.

In March 1999 Volvo Car Corporation passed into the ownership of the Ford Motor Company. Within Ford Motor Company Volvo Cars together with Jaguar, Land Rover, Lincoln and Aston Martin form the Premier Automotive Group, where Volvo Car Corporation is "Center of Excellence for Safety" (Volvo, 2004).

The unit for Driver Information & Interaction Design at Volvo Torslanda in Gothenburg, Sweden, is one department at the Volvo Car Corporation, striving to conform to the human focus by enhancing usability.

2 RESEARCH QUESTION

The purpose of voice control in a vehicle was, at first, to let the driver use his/her hands and eyes merely for the driving task and in so doing enhancing traffic safety. But letting the driver use his voice instead of his hands to operate secondary functions, such as turning up the volume on the radio, does not necessarily mean that the mental work load always decreases. Some voice control systems (see Lexicon) are complicated to operate and do not minimize the workload for the driver, which means traffic safety is compromised.

The questions we raise in this thesis are:

1. *How can an improvement plan to the voice control prototype operating the navigation system in the Volvo XC90 be made so that usability and safety are at focus?*

2. *Could multimodality be implemented in the Speech User Interface for this purpose?*

This master thesis will contain an evaluation and a user study of a voice control prototype within the navigation system for a Volvo XC90. It will also contain an evaluation of the voice control for Jaguar Advanced Voice Portal (AVP). With the results from the user study and evaluations an improvement plan will be made.

The work will include:

- A literature study
- A performance study; a task oriented user study of the voice control in the XC90.
- An attitude study: What does the user think of the Volvo Voice Control Prototype; a subjective point of view?
- A cognitive walkthrough of the Jaguar Advanced Voice Portal (AVP).
- Design of an improvement plan.

2.1 DELIMITATIONS

The voice control system that has been tested is a prototype installed in a Volvo XC90 with Volvo's navigation system and does not include other parts of the infotainment system.

The improvement plan does not include improvement suggestions for the Graphical User Interface (GUI) (see Lexicon) of Volvo's navigation system.

Only the driver's situation has been observed, the passengers situation is not included in this study.

In the improvement plan a design proposal will be given but no prototype will be made.

3 THEORETICAL FRAMEWORK

In this chapter a theoretical framework supporting the research question is outlined. As this thesis is about speech technology, this area is described to enable the understanding of Speech User Interfaces. Speech User Interfaces can be classified according to certain variables which are explained to better understand the characteristics of the system that is evaluated. A part in this thesis work is to make an improvement plan for the voice control prototype, thus Speech User Interface guidelines has been studied and used to identify positive and negative aspects of the current system. The set of guidelines are to be found in Appendix A.

The Speech User Interface at focus is designed to operate the navigation system in the Volvo XC90. Hence to understand the essentials in navigation, and how humans do to navigate, some research on human navigation is presented in this chapter. This knowledge was later used to make a task analysis. In this part of the theoretical framework, in-vehicle navigation systems are briefly described and Volvo's navigation system is introduced.

The second research question examines whether and how to allow multimodality to operate the navigation system when making the improvement plan. Some general reasons for implementing multimodality are listed followed by an analysis of what modality (auditory or visual) is appropriate to present different types of information.

Lastly, essentials in human capacity for handling different types of information are explained and related to the driver environment.

3.1 SPEECH RECOGNITION

Using a Speech User Interface is a way to make the communication between humans and computers more similar to the communication between humans. A Speech User Interface is an interface allowing speech as input and/or output. The technology used for speech input is called speech recognition.

Speech recognition enables computers to identify human speech out of an environment with a variety of sounds and use it as input to control a system. First attempts with automatic speech recognition were made during the 1940's when telephone companies tried to construct a machine for recognition of numbers with the aim to replace telephone operators. The machine was at that time a challenge too big for the technology and another solution was used (Blomberg, Elenius, 2000; Hammar, 1995). During the last years the technology has improved and there's now a possibility to use automatic speech recognition in commercial systems.

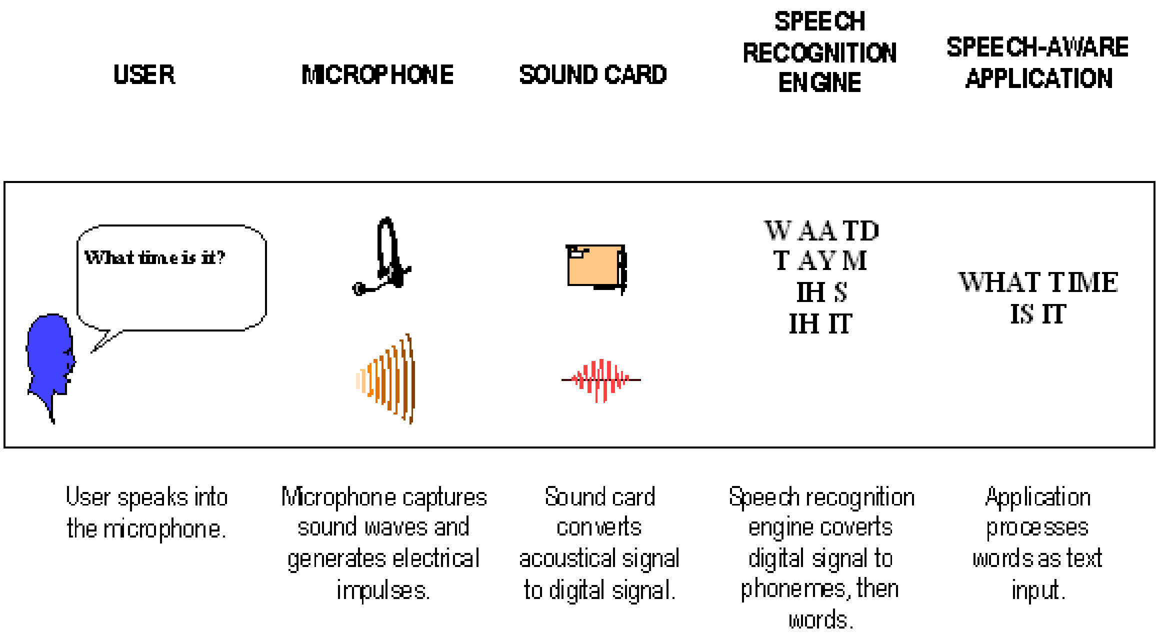


Figure 2: Shows the speech recognition process flow (Microsoft, 2004)

When human speech is used as input to a computer the spoken utterance, which is built up by sound waves, is captured by a microphone which generates electrical impulses. The acoustical signal is converted into a digital signal by the sound card and then the speech recognition engine converts the signal into phonemes and then into words. Finally the text input is processed by a speech aware application.

3.1.1 Two types of speech recognition

Speech recognition systems can be divided into two categories: continuous and discrete.

Continuous recognition systems – enables the user to speak to the system in an everyday manner without being constrained to use predefined commands. The systems tend to be error prone due to the complexity in understanding phonology, grammar and semantics found in natural language. This type of recognition is expensive to develop and primarily found in medical and defense systems and in up-to-date dictation systems (Weinschenk, 2000).

Discrete recognition systems – has a predetermined lexical database with all the necessary commands. These commands are directly mapped to predefined actions in the system, thus the system only identifies the phonology in the speaker's utterance not using any interpretation of the grammatical or syntactical structure. A discrete recognition system identifies utterances without interpreting them as a continuous system does. The recognition rate is usually high in discrete systems. When speaking to a discrete system a pause is required between utterances. However there are discrete recognition systems using the *word spotting technique* which enables the user to utter an entire phrase which gives an illusion of the system understanding continuous speech. The phrase is filtered for commands and sorts out the rest of the utterance. The technique is also called *parsing* and allows the interaction with the system to mimic the interaction between humans (Weinschenk, 2000).

3.1.2 Speaker dependent and speaker independent systems

Automatic speech recognition systems can also be divided into speaker dependent and speaker independent systems. Systems belonging to the first category have to be trained before they reach their top level of recognition. The user reads a predefined text and the system optimizes the recognition to the user's voice and speech pattern. The accuracy in recognition is generally greater in a speaker dependent system than in an independent system but it can't be used by multiple users. Speaker independent systems can be used by multiple users and don't have to be adapted to a certain speaker. As a consequence, these systems are not as precise as trained systems and they are not suitable for tasks requiring a large or specialized lexicon (Weinschenk, 2000).

3.1.3 Speaking to a system

Uninterrupted capture and processing of speech, demands a good recognition engine and a lot of computational power. By allowing the recognition engine to be started and halted for the speech that is relevant the required computational power can be limited.

PTT-button

To enable input to a system a Push To Talk (PTT)-button can be used. The user pushes the button to activate the system's recognition engine. The recognition engine then listens for speech. Using a PTT-button is common in vehicles and other noisy environments. There are different ways to use a PTT-button:

- The user presses the button continuously while speaking, comparable to a walkie-talkie. An advantage is that the user knows exactly when the system is listening. Literature in ergonomics recommends avoiding pressing things while driving as it could conflict with the steering (Karlsson, 2003).

- The PTT-button has to be pushed before each utterance which gives the user maximum control over the system as he/she always initiates the listening. After a predefined time the listening ends.

- The PTT-button is used like an on/off-button and is pushed to start and stop the listening. This style of use involves high frequency of pushing. User studies have shown that users often forget to inactivate the listening (Karlsson, 2003).

- The user pushes the button to initiate the listening and the listening is automatically reactivated when the system has finished playing a prompt. After a certain time the listening is timed out. Compared to other ways of using a PTT-button this way may implicate difficulties for the user to know whether the system is listening or not.

Command activation

Another way to address a system is by using a spoken command to activate the recognition engine. The system then has to listen continuously for the start command and filter out other words and background noise something that requires a good recognition engine and constant computational processing.

Barge-in

Some systems allow the user to *barge in* which means that the speech recognition system can identify spoken commands while the system is playing a prompt (Brems et al, 1995). Using barge-in speeds up the interaction and makes the dialogue run smoothly as the user doesn't have to finish listening to prompts if he/she knows what to say. In every day speaking situations people often interrupt each other, which isn't always considered impolite. To allow barge-in in a dialogue system would thereby make the dialogue between the user and the system more similar to the one between humans witch results in a more natural dialogue. Together with the improved dialogue flow, the more natural dialogue is claimed by Heins et al (1997) to increase the overall user satisfaction with the system.

3.2 SPEECH SYNTHESIS

The first machine built to output human like speech was OVE (Orator Verbalis Electicus), a machine built by Gunnar Fant during the 1950’s (Fant, 2000). The machine could produce vocals and some simple phrases, it could also pronounce the Swedish name Ove, whereby it got its name.

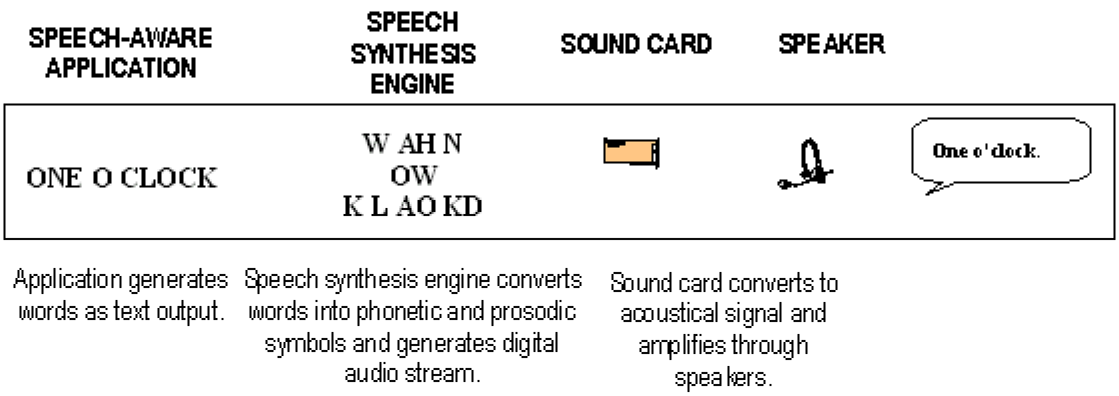


Figure 3: Shows the Text-to-speech process flow (Microsoft, 2004)

Speech synthesis is also called text-to-speech as it produces speech from a text. The first step in the process is to generate a text output. The words in the text are translated into phonetic and prosodic entities that can be transformed into a digital audio stream. The digital sound is converted into an analog signal emitted by the loud speakers (Microsoft). The voices in the system can be either digital or recorded from a human. To convert the phonetic and prosodic representation into speech there are two different ways: formant synthesis and concatenated synthesis.

3.2.1 Format synthesis

Format synthesis uses a mathematical model of the human vocal tract and complex phonological rules to create intelligible but unnatural sounding speech. The speech is entirely machine generated (a manipulated audio waveform). The advantage with this type of system is its ability to produce almost unlimited speech. Another advantage is that there’s no need of large storing capacity. A weakness with formant synthesis generated speech is that the unnatural sounding voice can annoy the users. The digital voice can also contribute to unrealistic expectations of the system potential.

3.2.2 Concatenated synthesis

Concatenated synthesis uses small segments of prerecorded human speech which is why it sounds more human-like than formant synthesis. The recorded units are assembled on demand according to given rules. The fact that the speech is recorded from a person makes the system less flexible than a system with formant synthesis when changes have to be made. In that case the same speaker has to record new segments. Concatenated speech synthesis is best used in systems with a small vocabulary as it requires numerous speech units, large storage space and a lot of computational power (Weinschenk, 2000).

3.2.3 Characteristics of good speech synthesis

Two factors characterize a good speech synthesis; **intelligibility** and **naturalness**. Intelligibility refers to in what degree the user recognizes the phrases and words used by a system. If there is a high level of intelligibility the users will perceive the system as being very precise and they will, in general, be satisfied with the auditory experience. It is possible to reach almost the same level of intelligibility from a system as it is from a human. The level of accuracy in understanding synthesized speech can be measured by comparing the number of phonemes produced by a system and the number of phonemes correctly interpreted by users (Weinshenk, 2000).

As many societies in the world are increasingly multilingual, especially in Europe, the need for text-to-speech devices that have multilingual capabilities increases (Roe, Wilpon, 1994). By using a TTS that is multilingual the system will increase the level of intelligibility.

Naturalness refers to how similar the synthesized speech is to human speech. It is a subjective measure and varies with the users, the task and the context. It is not necessarily something desired. In an airplane's cockpit for example a less-human sounding voice could be preferred to avoid that it is mistaken for radio communication (Weinschenk, 2000). Another reason to have a less human-like sounding voice is to avoid making the user believe that the speech recognition has no functional and technical limitations and addresses it as if it were human, that is; with natural speech.

3.3 NAVIGATION

Humans navigate through spaces frequently throughout a day, sometimes in known areas and sometimes in unknown. To navigate we have to know where we are, the current position, and where we are going, the destination. The knowledge of the latter can be based on experience if the destination has already been visited. If the destination is not previously known the mental image is created from other known places; an expectation of what the place might look like. To reach the destination a plan has to be made and there must be a transportation of the person or object. This is described by Gallistel (1990) as follows:

> *Navigation is the process of determining and maintaining a course or trajectory from one place to another. Processes for estimating one's position with respect to known world are fundamental to it. The known world is composed of surfaces whose locations relative to one another are represented on a map.*

The map mentioned contains information about the world and exists in the brain. The map is not an exact replica of reality but a simplified image that represent the person's expectations. How it's represented and implemented is explained in various ways in studies in ethology, psychology, artificial intelligence and neuroscience.

Levitt and Lawton (1990) describes navigation with three questions:

1. Where am I?
2. Where are other places relatively to me?
3. How do I get to other places from here?

The answer to the first question does not have to be an exact position on a coordinate system but may be constituted of characteristics of the place. To answer the second question, a global spatial representation which represents every other position related to the current position in a spatial relation network, can be made. The answer to the last question includes planning the path to the goal with all its waypoints.

3.3.1 In-vehicle navigation systems

Many car brands offer navigation systems as an option in their cars. These systems do not only calculate routes and provide the driver with driving instructions but also hold additional information about the driver's environment such as traffic and road information and public establishments such as restaurants and post offices.

Most common input devices used to interact with in-vehicle navigation systems are buttons, haptic devices (see Lexicon) or touch screens. Some systems can also be operated by voice control.

The output from the system can be both visual and auditory; often the user can choose to simply get guiding instructions visually on a display, or reinforced with voice guiding.

3.3.2 Volvo's navigation system

Volvo started to develop the RTI (Road Traffic Information) system in 1996 and the navigation system is today optional in all Volvo cars. It is designed to provide the driver with route planning, driver guidance (both visually and auditory) and traffic information. The system has a display situated on top of the dashboard that can be lowered down into the dashboard and out of sight. Input to the system is given by using three integrated buttons (Back, Enter and a four-way switch) on the back of the steering wheel or by using a remote control. The navigation system has prerecorded voice guiding instructions and written text available in English, French, German, Spanish, Italian, Portuguese, Dutch and Swedish.

Figure 4: Shows the navigation display (left), the *Back*- and *Enter*-buttons and the four-way switch on the underside of the steering wheel (right).

To keep track of the car's position GPS technology is used along with a navigable database containing addresses and digital maps. This information is stored on a disc and upgraded on a regular basis.

3.4 VOLVO VOICE CONTROL

As mentioned in the introduction (see Introduction) Volvo has developed a voice control prototype for the navigation system.

3.4.1 The prototype of today

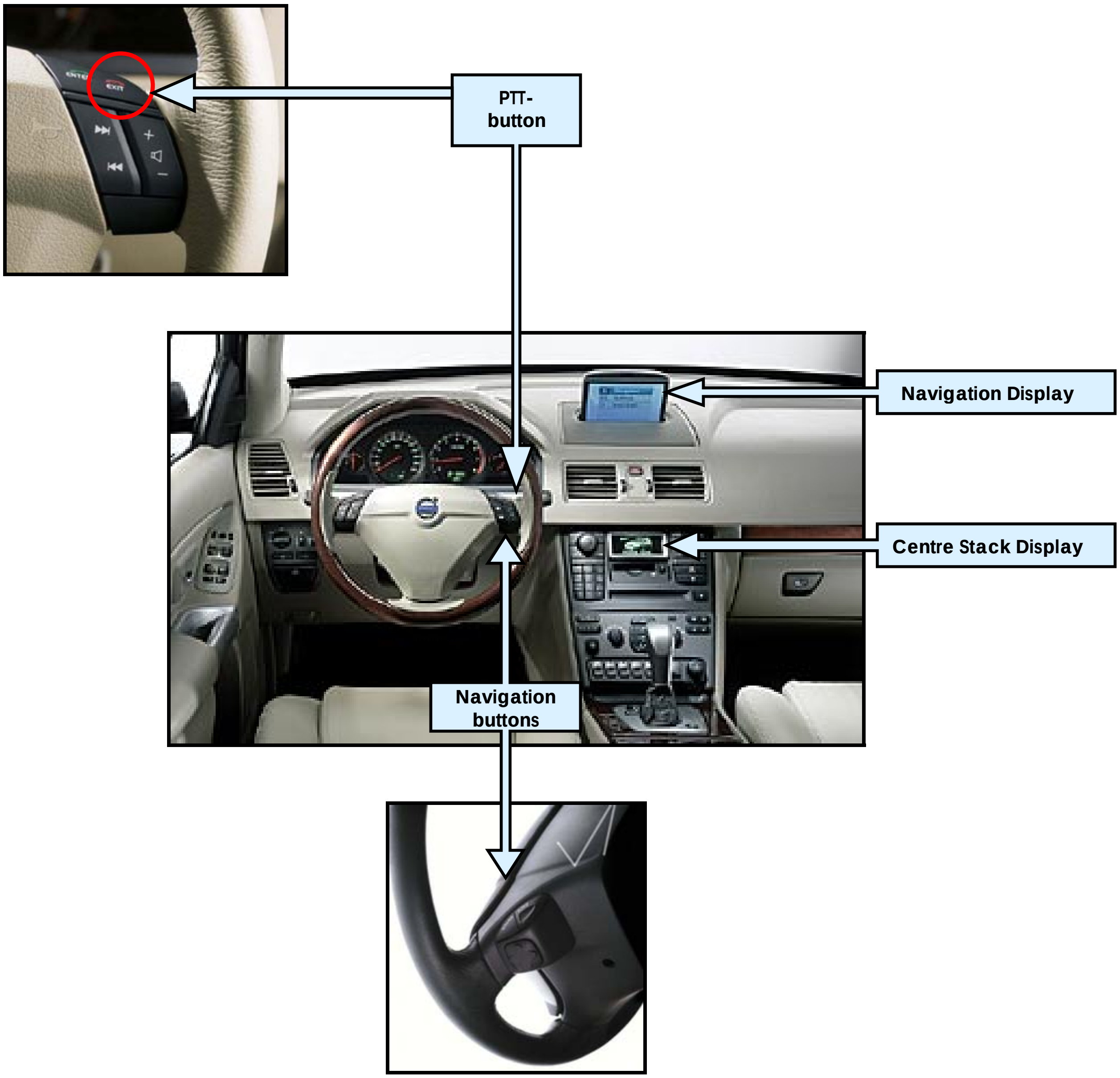


Figure 5: Shows the interior of the Volvo XC90 with Volvo's navigation system and voice control.

A brief description of the Volvo Voice Control Prototype

- The system uses discrete recognition and recognizes about 200 commands.
- English is the only language implemented.
- The system is speaker independent.
- The system is designed for the driver, excluding the passengers to use it.
- A PTT-button on the steering wheel is pushed before each utterance to start the recognition engine. When the PTT-button is pushed the system waits until a command is uttered by the user. If no command is spoken the recognition engine is timed out after 4 seconds. The PTT-button is also used to barge in.
- The system's help function is entered by uttering the command *What Can I Say (WCIS)*.
- The help function (*WCIS*) is context aware (see Lexicon).
- Feedback is given both visually and auditory. The system outputs speech, non spoken sounds and graphics.
- The Text-To-Speech (TTS) use concatenated synthesis to produce English speech.

Help menu structure

When the user starts to interact with the system it remains silent until the user gives a command. On condition that the command is recognized the requested action is performed or a dialogue is initiated. If the command is not properly recognized, the system informs the user about this. Instructions on how to access the help menu (by saying "*What Can I Say*") are not introduced to the user by the system. The help menu is auditory and informs of most commands in the system. As mentioned the help menu is context aware. The information provided is depending on whether there is an ongoing task with voice control or not. When no specific task is in progress the help menu available to the user is divided into three main categories which are described below.

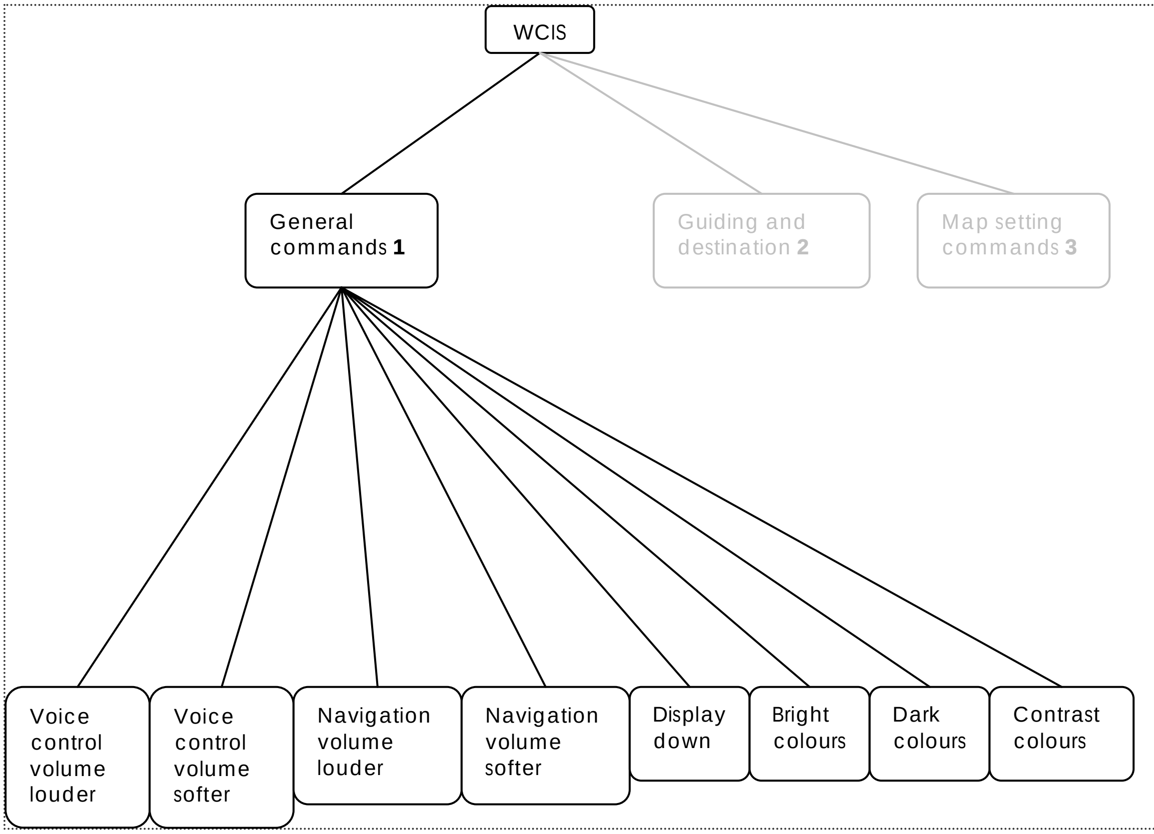


Figure 6: Shows the category *General Commands* in the help menu

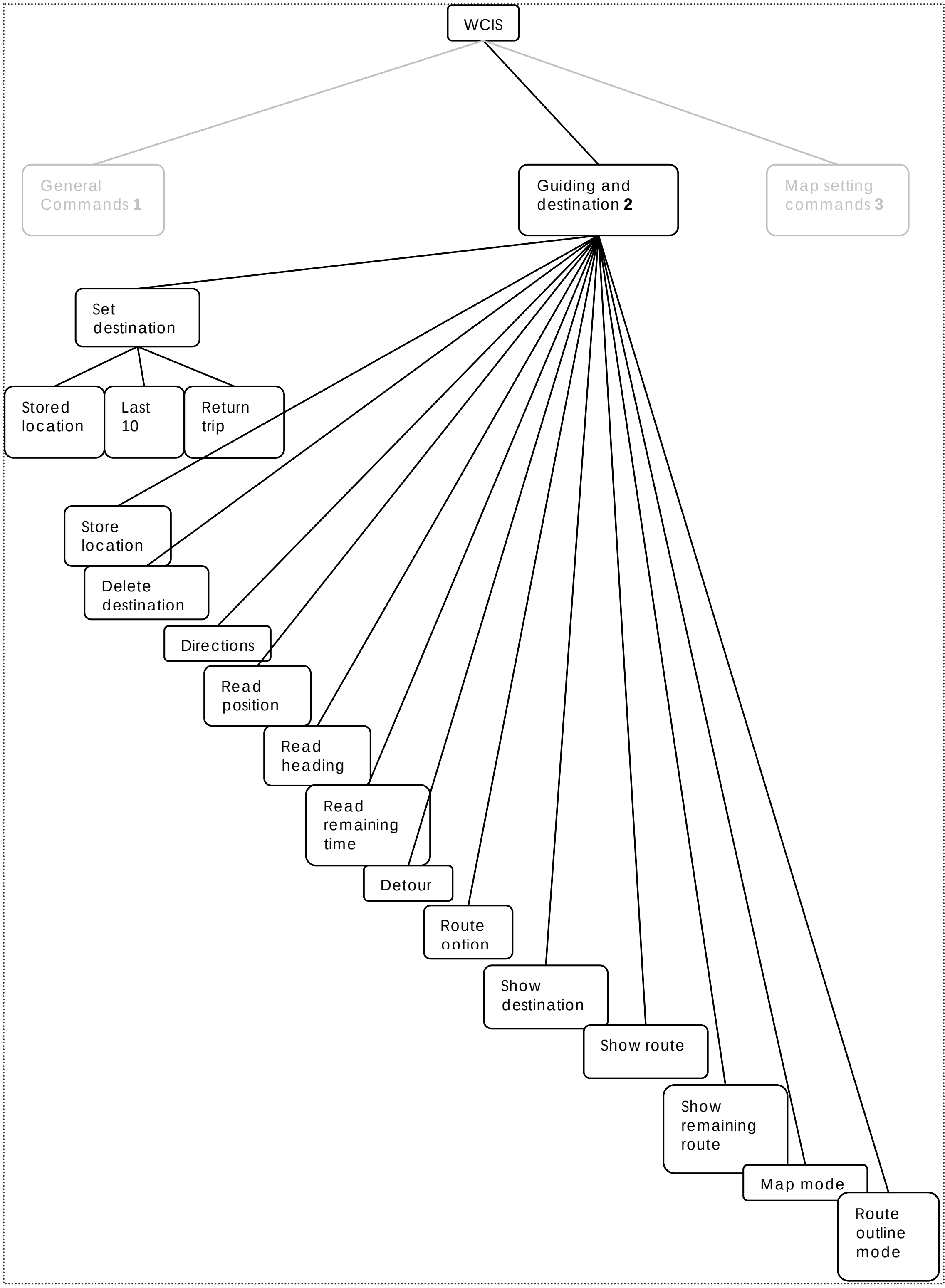


Figure 7: Shows the category *Guiding and destination* in the help menu.

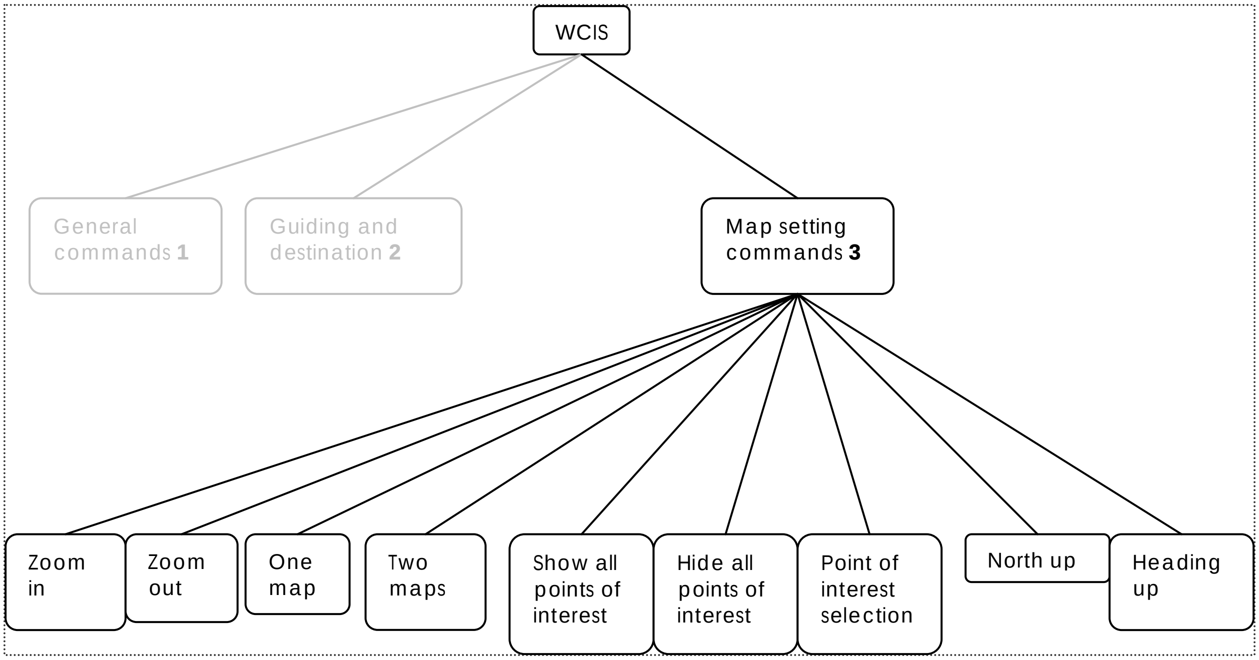


Figure 8: Shows the category *Map setting commands* in the help menu.

3.4.2 The history of the prototype

During the development of the Volvo Voice Control Prototype three usability studies were performed. The information about the tests has been gathered from internal Volvo documents. The performance of the speech recognition engine was not evaluated during these tests. The two first tests were conducted on a prototype that was not installed in a vehicle.

User test 1

The first test was a small-scale study conducted in a desktop environment at Volvo Torslanda, Sweden, with Volvo employees as test participants. The voice control prototype was not installed in a car, thus the test used two PC desktop simulations. Ten participants aged between 25-65 years took part in the test, 50% of them were women and 50% of them men. The purpose with the first study was to find the major problems with the system and fix them before the second, and more extensive study, was conducted in the USA.

The aim with the first test was to investigate if the users knew:

- Where they were in the system.
- What to say.
- How to express it.
- If they were heard correctly.
- How to fix errors.

The tasks concerning the navigation were to:

- Set a new destination.
- Change the display view to two maps, and then back to one map.
- Zoom in the left part of the display one step.

All test persons managed to perform all the tasks. Two general usability problems were discovered. The first was that the auditory help was found too long and the users found it tiresome and frustrating to listen through the same lists several times. The second was that many users did not use the auditory feedback (a ping sound) given by the system when the PTT-button was pushed. Not all of the users were sure of what it meant (the ping sound indicates that the system starts to listen).

User test 2

The second user test took place in Tustin in California, USA. The aim with this study was as well to test usability but this time the test participants' mother tongue was American English, thus one source of error was eliminated. The test tested changes made after the first usability test.

The study consisted of two parts; one-to-one interviews to discover usability problems and focus groups to get a deeper insight into the respondents' attitudes and ideas.

Thirty test participants participated in the one-to-one interviews and twelve in the focus groups. They were aged between 25-55 years, and as in the first test, half of them were men and half of them women.

The one-to-one interviews were conducted in a desktop environment and used the same simulations as used in the first test but with the changes made from the results of the first test. This time the participants performed the tasks already described but also additional tasks: to make the system read out the time left to the destination, to store a destination, to show all traffic information on the map, delete all last destinations and make the system read out information on an icon. After performing the tasks the participants filled in a questionnaire and a short interview was held. Some of the participants later took part in one of the two focus groups.

The results of the usability test showed that the users wanted a back or undo-function similar to the manual one. Several test participants found the messages in the help menu too long. They would rather see a deeper menu hierarchy with fewer commands. To initiate the help function the word help was preferred to *What Can I Say*. Some of the users said "What should I say" instead. The participants had difficulties to get the system read out the remaining time to destination. They used a broad variety of alternative commands. However, the command *read remaining time* was recommended to be kept as the same vocabulary is used in the Graphical User Interface.

People wanted to be able to change category in the help function without being obliged to repeat "What Can I Say" if they had entered the wrong category.

Open-ended questions tricked people to say the wrong thing. Instead of specifying the action sought for (such as show or hide), people answered *yes* to questions like "Do you want to show or hide all Points of interests?".

User test 3

In 2003 a study of the users' attitude towards voice control was conducted, this was the first test of a prototype in a car. The total number of participants was 21 of which 13 were men and 8 women. The participants' average age was 34 years.

The test was designed so that the participants tried out the system after a short introduction from the test leader. No test leader was at hand during the test and no observation of the drivers was made. After having experienced the voice control the participants filled in a questionnaire with questions about:

- system comfort while driving
- the systems recognition rate
- the choice on what functions to control by voice
- the system's effect on driver safety
- estimated mental workload (see Lexicon)

To answer the questions a ranking between 1 and 5 was chosen. The result showed that on the whole participants were satisfied with the system but that they experienced it to affect the driving and the mental workload.

Comments on the system were collected. Some positive remarks talked about the simplicity in using voice to operate the navigation system, some considered voice interaction to be easier than manual interaction. The users contributed with ideas on how to improve the system, some of these suggestions are listed below.

- Show auditory menus graphically as well.
- Allow alternative commands.
- Reorganize the menus in the help function.
- Allow setting a new destination.
- A better TTS.

3.5 MULTIMODALITY

Multimodal systems use several modalities to present information to the users, just as it’s possible for the user to interact with the system by using different modalities. In this report when referring to multimodal systems, systems accepting both speech and manual input and presenting the output both in graphics and speech are intended. The idea is to let the user decide when to use what modality and also to let the user change modalities within a task. There are however many different opinions on the definition of multimodality.

Human communication is by nature multimodal. There is actually a very small part of a message that is communicated by actual words between people (see Introduction). Most of the message is communicated by means of gestures, mimics and nonlinguistic sounds (laughter, clearing the throat, whistling…). It also seems as people naturally exploit all the available communicative resources and communicate according to the
Multimax Principle (Bunt, 1998)**:**

> *In natural communication, the participants use all the modalities and media that are available in the communicative situation.*

To see examples on how different modalities can be combined see Modality Combinations in Lexicon.

3.5.1 Reasons for applying multimodality

The second research question in this thesis deals with the issue whether multimodality can be applied to enhance usability for the Speech User Interface and thereby increase driving safety.

> *Many tasks lend themselves to multimodality. For example, a traveler may point to two locations on a map while saying “How far”. People will use one modality (such as speech alone) followed by another modality (such as pointing on a map with a mouse or a pen). In other words they will switch between modes. Sometimes they use two or more modes simultaneously, or nearly so, for example, pointing first then talking (Weinschenk, 2000).*

This describes a spontaneous way of interacting with something. Allowing multimodality for operating a system might facilitate the interaction and give the user the feeling of being in control. The user is more free to decide how he wants’ to communicate his message.

Using multimodality does not only support the human in providing systems with input. If for example spoken output is enhanced by graphic output the fact that auditory information is temporary is of less importance and the advantages by both types of information can be derived.

> *Because a speech-only system lacks visual feedback or confirmation it is taxing on human memory. Long menus in telephony applications, for instance, are hard to remember (Weinschenk, 2000).*

Multimodality could be used to speed up the interaction by letting the user choose how to interact with the system. Some parts of a task can be more quickly performed by using voice respectively a manual device. It is important to use the right modality for each task. Manual interaction is suitable e.g. for scrolling, clicking, pointing or choosing from a list. Speech interaction can be used to select available visual objects or to request a certain action to be performed.

According to Weinschenk (2000) auditory presentation of information (output) is effective when:

- information is short and simple
- information is needed immediately, but not later on (does not have to be remembered)
- information is temporal in nature (refers to events over time)
- the message is a critical warning
- a verbal response is required
- the visual system of the person is already overextended
- the environment is not conductive to visual display (for example, if the lightning is insufficient)
- the person need to stay "dark adapted" (that is, it is important that light levels stay low)
- the person needs to be moving constantly

Visual presentation of information is effective when:

- information is complex and long
- information needs to be remembered
- information deals with spatial relationships; for instance, maps
- the person's audition is overextended
- the environment is noisy, or creating noise in the environment is not acceptable

3.5.2 Memory

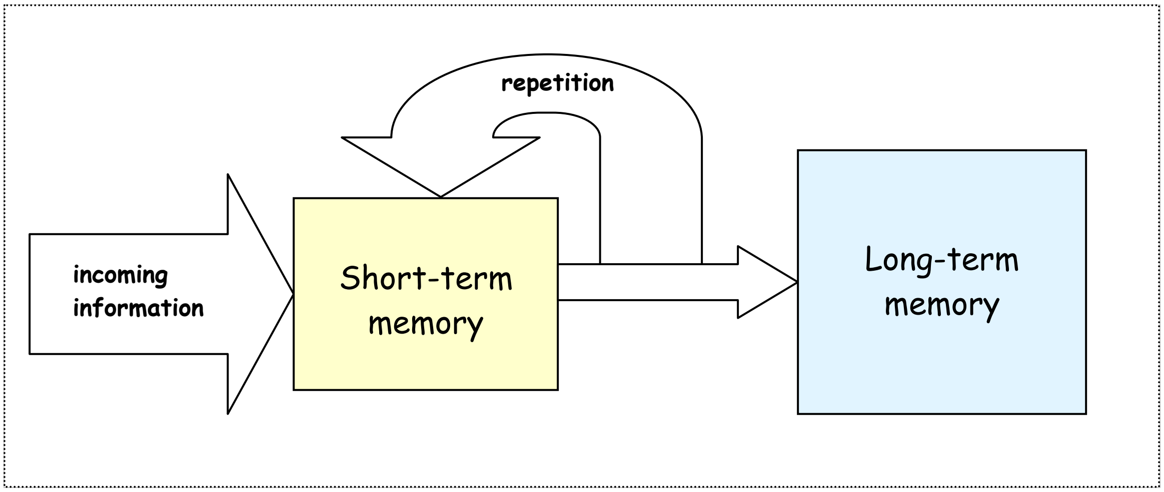


Figure 9: Shows the relationship between short-term and long-term memory (Lundh et al, 1992).

Information is temporarily stored in the short-term memory. This memory is also referred to as *working memory* as information that is currently being processed is kept here. When the information is no longer being processed it usually vanishes (Lund, Montgomery, Waern, 1992). The short-term memory is very unstable and can lose information if disrupted. The average human can keep 7±2 *meaningful* units of information in the short-term memory at once (Wickens, 1992).

Information that is repeatedly processed and stored in the short-term memory can go on to be stored in the long-term memory. Information that is stored in the long-term memory can be retrieved from here and used in the short-term memory for example to solve a problem (Lund, Montgomery, Waern, 1992).

3.5.3 Mental Multiple-Resources

It can be questioned if there is only one single supply of undifferentiated mental resources and the multiple-resource view argues that people have several different mental capacities with resource properties (Wickens, 1992). Also see resource capacity in Lexicon.

Resources can be defined by three dichotomous (see Lexicon) dimensions (Figure 10): two stage-defined resources (early versus late processes like encoding and responding), two modality-defined resources (auditory versus visual encoding), and two resources defined by processing codes (spatial versus verbal).

This means that information retained in the working memory can be represented by two forms of code: spatial, e.g. visual images, or verbal, e.g. written words and spoken words.

According to Polson & Friedman (1988) spatial and verbal processes (*codes)*, whether functioning in perception, working memory, or response, depend on separate resources (Grane, Rydström, 2002). They seem to work independently from one another. Our thesis will specifically focus on a driver situation, where the operator (driver) is presented with a task environment imposing heavy demands on spatial working memory. Due to the heavy demands on the spatial working memory, reserve capacity will be most likely to be found within the verbal working memory. Wickens (1992) recommends that, in for example a driving situation, verbal tasks are mostly *compatibly* responded to with voice control and spatial tasks with manual control. This is illustrated in the picture below (Figure 10).

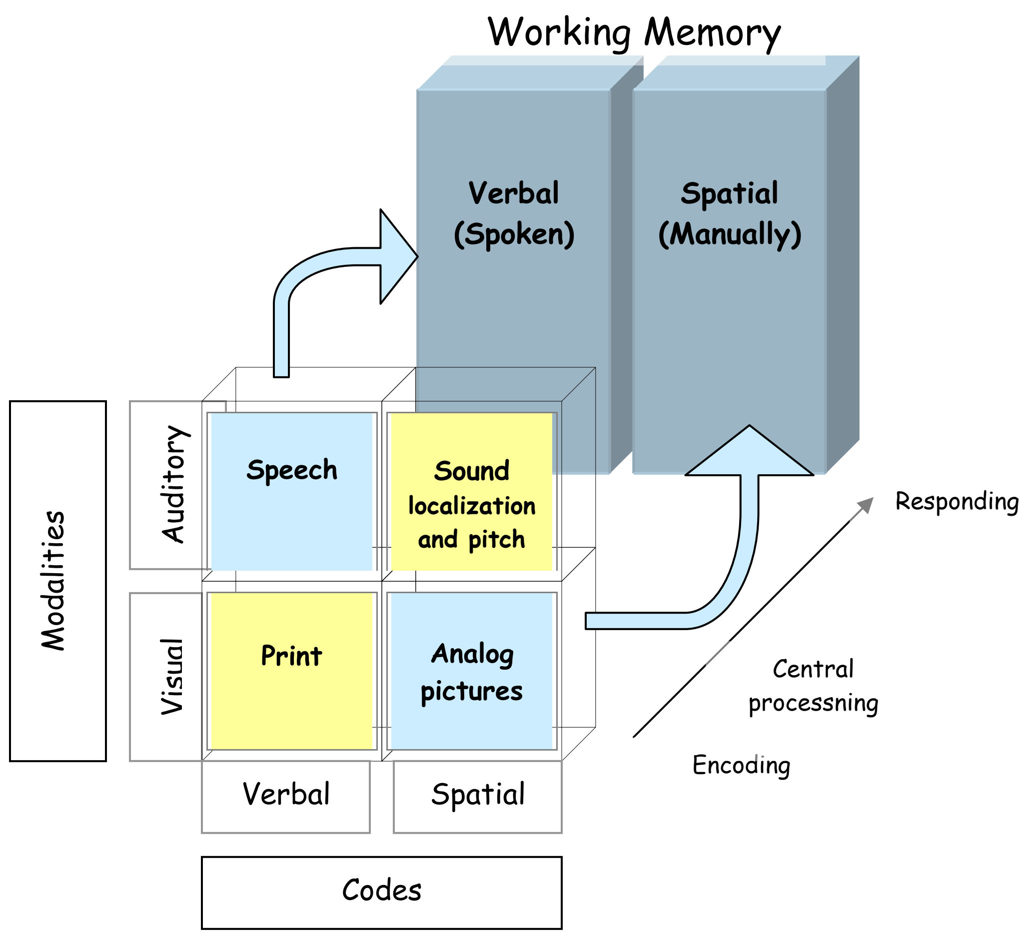


Figure 10: Shows the optimum assignment of display format to working memory code ((Wickens, 1992) and parts of picture (Grane, Rydström, 2002)).

Information can be displayed in four different formats: by speech, by sound localization and pitch, by print and by analog pictures. This information can be displayed in two modalities: visual (*print* and *analog pictures*) and auditory (e.g. *speech* and *sound localization and pitch*). Examples of print are written words or numbers. To make sense out of written words the language center of the brain is used in the same way as when spoken words are used. The two resources defined by processing codes (spatial versus verbal) represent how the information is processed in the mind, and what areas that are used to help solve a problem.

The verbal mental memory uses the language center in the brain to make sense out of a written or spoken word (verbal code). The spatial mental memory uses other parts of the mind to make sense out of, for example, analog pictures (spatial code). An analog picture is a direct depiction of something, like the picture of a car. To understand the picture and make sense of it the long term memory is used where other mental images are revised to search for similarity. This means that the picture of the car does not need to be exactly the same as other cars seen before, as long as it has the characteristic features of a car.
Finally, the two stage-defined resources (early versus late processes) are represented in the picture as encoding, central processing (early), and responding (late).

4 METHOD

4.1 INFORMATION RESEARCH

During the beginning of the research a literature study of the speech technology domain and speech utilities in vehicles was carried. Continuously literature has been consulted to learn more about disciplines considered in this work.

Speech technology has developed and highly improved during the last years, therefore the most recent information is often found on the web.
Internet sites of different car manufactures were visited to learn about how speech technology is applied in a car environment today and what companies use speech technology to interact with navigation systems.

In addition to literature and web research telephone conversations were held with car dealers about the use of speech technology in these car brands. There was also an opportunity to briefly review the Denso Speech User Interface for operating the Denso navigation system. At a fair, representatives from BMW were consulted about the company's navigation system and speech technology system.

4.2 GUIDELINES FOR SPEECH USER INTERFACES

To make an evaluation of Volvo's and Jaguar's Speech User Interfaces, a set of speech guidelines have been used as a reference. These guidelines were chosen as they are adapted to Speech User Interfaces and derived from 20 laws of interface design. The laws are based on human factors issues and are a result of a comparison of the work of Ben Shneiderman (1998) and Jakob Nielsen (1994) made by Weinschenk and Barker (2000). Weinschenk and Barker have also contributed to the set of laws as well as Julie Nowicki (Optavia Corporation). Some of the guidelines were excluded due to the fact that they are specific for speech systems using continuous listening which is not applied in the systems to be evaluated. The guidelines can be found in Appendix A.

4.3 USER CENTRED DESIGN

As the name indicates, user requirements and user needs are at focus in user centred design. The method is iterative and includes four main stages.

Figure 11: Shows one level in the iterative process (MDI, 2002).

1. Analysis – a user analysis and a task analysis is outlined
2. Requirement Specification – usability demands and functionality demands are identified
3. Design and prototyping – design derived from the requirement specification
4. Evaluation – user testing based on the requirement specification

These stages are iterated until the design matches up sufficiently towards the requirement specification.

In this thesis the first, second and the last step has been considered. As the aim of this thesis was to make an evaluation of an existing prototype, the four steps have already been iterated several times before. However, to create our own image of the users and the task both a **user analysis** and a **task analysis** were made.

To make the user analysis the main market for the Volvo XC90 (the type of car into which the voice control prototype is built) was looked at. A user analysis gathers a broad variety of information about the users and the user environment. After having identified the users, qualitative interviews were held with 4 possible users to get a more complete image of their habits, needs, interests and so on. The interviews were semi-structured with both open and closed questions. The interviewees were allowed in part to steer the conversation, bringing the interview into unexpected turning. To conclude the interview the interviewees answered a multiple choice question about their attitude towards new technology. Questions and answers are found in Appendix B.

Volvo's existing navigation system has an extensive list of functions and to find out the core functions to be operated with voice, the task analysis was made. The task analysis describes the users' goals, the characteristics of the task, how the task is performed today and the advantages and disadvantages with the current solutions.

The next step was to write scenarios (Appendix C) to illustrate the image of the user and possible user situations. Out of the result from the user analysis and the task analysis a list of usability demands and functional demands was made.

4.4 JAGUAR TEST

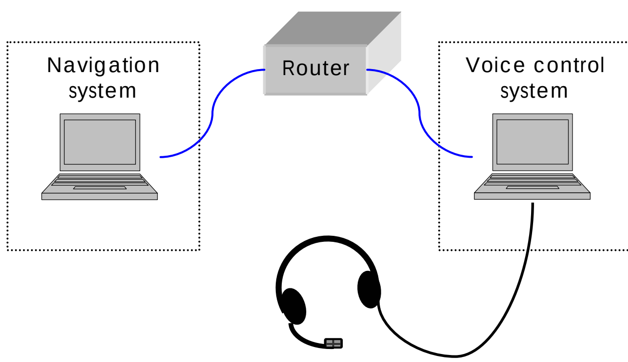


Figure 12: Shows the setup used during the cognitive walkthrough of the Jaguar AVP.

Due to the fact that the Jaguar AVP prototype was not implemented in a car, it was not possible to conduct any user test comparable to the user test of the Volvo Voice Control Prototype. The Jaguar AVP prototype was too incomplete to give valuable results from a user test with real users. Instead an expert evaluation of the system was performed.

The Jaguar prototype was installed on two laptops and the expert evaluation took place in a somewhat silent room and not in the future system's real context; that is in a car.

In a cognitive walkthrough a number of usability experts make a walkthrough of a system. A set of predefined tasks are to be completed within the system and each expert performs these tasks independently from the others. The usability experts enter the role as representative users of the system and look for possible actions that can lead them towards the goal. They also search the system for possible difficulties which can cause breakdowns in the interaction. The aim with this type of evaluation is to answer the questions: will users know what to do, how to do it and will they understand the feedback given by the system? To find the system's positive and negative aspects, it is important that the expert really tries to act like a user and that the expert has a correct image about the users' skills, limitations and probable behavior (Faulkner, 2000).

The tasks in the expert evaluation were generally the same as the tasks for the user test of the Volvo Voice Control Prototype but adapted to the Jaguar AVP prototype. As opposed to the user test in the Volvo, there were no comparison of manual and speech interaction in the expert evaluation as the Jaguar prototype only could be operated by speech. Both input and output was auditory in the Jaguar prototype as no graphical interface was available.

A brief description of the **Jaguar AVP prototype:**

- The prototype was installed on two laptops; one with voice technology system, one with the navigation tool.
- The system uses an application specific dictionary e.g. the street names are stored phonetically on the navigation map disc.
- The system uses discrete recognition and recognizes about 150 fixed commands and approximately 2000 street names.
- The system is speaker independent.
- The system's functions are designed for the driver. To give input to the prototype a headset is used. In the future application the microphone will be fixed in the car and aimed at the driver.
- The *Enter*-button was used as a PTT-button. When the button is pushed the recognition engine is activated and the system addresses the user by reading out loud the choices and commands in the main menu. If there is no input to the system, the main menu choices are repeated two times and then the interaction is timed out.
- If the user wants to barge in while a prompt is played, the user is not required to push the PTT-button before saying a command. In the future application Jaguar has decided to let the user push the PTT-button each time before making an utterance (as in the Volvo Voice Control Prototype).
- The help function is called on by using the command *help.* On production, saying "*What Can I Say*" will also start the help function.
- Feedback is given auditory but in the production system feedback will also be given visually. The system outputs speech and non spoken sounds.
- The text-to-speech (TTS) use concatenated synthesis. The language output is in English.
- Prerecorded guiding instructions are available in English.

4.5 VOLVO TEST

4.5.1 Brainstorming

To be able to generate a vast amount of ideas on what aspects of the system that needed to be tested a brainstorming method was used in addition to the task analysis that was made earlier. All the questions that popped up during the session were written down on post-it notes and placed on a wall. The notes were then categorized into sections. Redundant notes were thrown away. Some of the remaining post-it notes that were either too general or too specific were derived to fit into an appropriate group or form a new one.

The results of the brainstorming session were three main categories; questions and tasks to be asked/performed before, during and after the test.

4.5.2 Observation methods

There are different techniques to test usability and one of them is ***Think aloud***. This technique requires the user to say out loud what they are trying to do, why they are doing something, and what they are thinking during the interaction. This means that as the user is interacting with a system the user also communicates what is on his mind as interaction occurs (Dumas, Redish, 1999).

Another method is **Co-discovery**. This involves having two participants work together during the test as two people talking to one another is more natural than to *think aloud* alone. As a result, co-discovery tests yield more information about what the users are thinking and what strategies they are using to solve the problems (Dumas, Redish, 1999).

The co-discovery group consisted of three pairs. The other participants were tested individually. The group with single participants performing the test generally asked for more help than the co-discovery group.

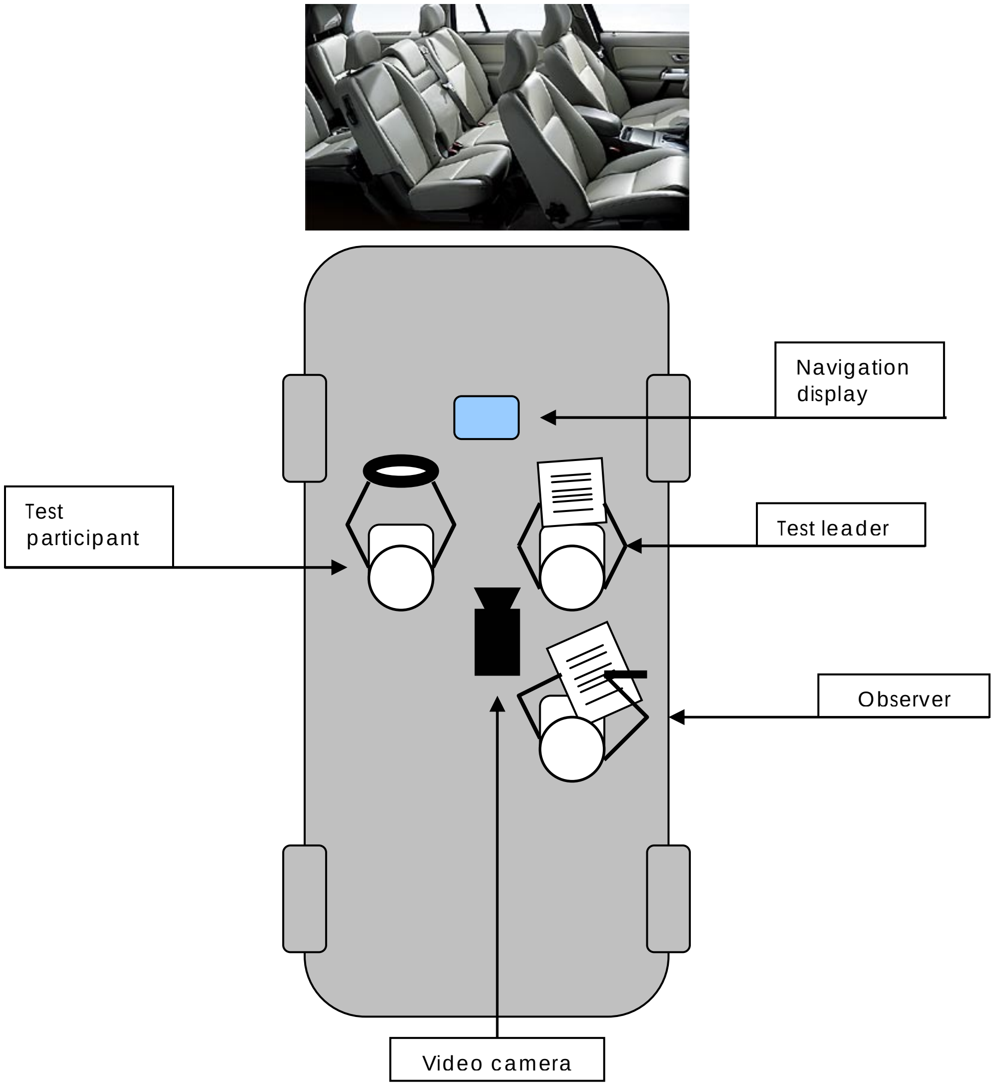


Figure 13: Shows the test situation in the Volvo XC90.

The test team roles

The test team was divided into two roles; one observing and the other conducting the test. The one observing the test was also the one in charge of the video equipment used in the test. These roles where switched halfway through the testing period, even though this is not generally recommended, this was done for training purpose and to prevent monotony. Monotony can toll on concentration and result in the observer not listening attentively and hence missing vital information.

The test leader was during the single tests seated in the front seat, and the observer in the back seat. All instructions about the tasks where given by the test leader and subsequently the test leader was the one answering any questions asked by the test participants. The observer remained mainly quiet in the background during the test. During the co-discovery stage the test leader and the observer were seated in the rear but still had the same roles and functions as during the single test.

Protocols used during testing

Not to miss any vital information observation protocols (Appendix G) were made. Two main kinds of observation protocols were used; one during the test and the other during the interview. One task observation protocol was used by the observation leader and another simpler one was used by the test leader. The same interview protocol was used by both the observer and the test leader.

Checklists

Checklists were made and used before, during and after the tests (Appendix E). The checklist consisted of reminders on everything from what to say during the introduction, handling the video equipment, to practicalities as notes to be taken at the end of every test such as recording the mileage.

Video observation

To be able to go back and analyze specific situations that occurred during testing the test sessions were video taped, but only after getting permission from the test participant. No facial features were caught on film; the camera was aimed at the display of the navigation system to catch manual interaction and graphical feedback. Speech was also recorded through the video camera to log a more detailed interaction with the SUI.

Video analysis focus protocol

A video analysis focus protocol was made to facilitate the analysis of the recorded material. This protocol was made to establish the focus of the video analysis; what major areas to study.

A total of 21 hours of recorded material was gathered. Analysis of video data can be very time consuming if every action and utterance is to be analysed and transcribed in detail. Instead the video recordings were used to look for patterns, breakdowns and major difficulties in the system. Interesting comments from the participants were also noted.

Video observation protocol

A video observation protocol was made to facilitate and structure the analysis of the recorded material; what events to look for. This protocol can be found in Appendix H. The recordings for each participant were scrutinized, starting with the test followed by the interview. The one who had filled in the test protocols during the test and the interview checked these so the notes were correct and also completed them with lacking information. The other person noted observations of interest from the recording of the test and filled in the video observation protocols from the recording of the interview. When this was finished the information in the test observation protocols and the video observation protocols were summarized.

4.5.3 Test environment

The test was conducted in a Volvo XC90. This vehicle was equipped with Volvo's navigation system with RTI (Road Traffic Information) described in the theoretical framework (see chapter 3.3.2) and a voice control prototype. A built in microphone positioned on the rear-view mirror was used.

The tests were conducted during daytime (from 9:00 am to 3:00 pm) in daylight, during December, under normal traffic conditions, free from snow, around *Volvo Torslanda in Gothenburg.* The user could initially choose which route to take and was then instructed to follow a route guide as much as feasibly possible.

4.5.4 Questionnaires

Each member of the test group was given a questionnaire (Appendix D) to determine the group's general characteristics, such as age, gender and driving abilities. This information was also used during the individual analysis of the test persons test performance.

4.5.5 Participants

The test participants were all employees at Volvo Car Corporation. Both men and women participated and the average age was 41 years. Most of them had participated in other tests before and were used to test situations. Initially the aim was to test the system on the target group, the group of users that were most likely to use the system in the future, that is; American stay-at-home-mothers. Since the tested system is not yet out on the market this could not be done out of business confidentiality.

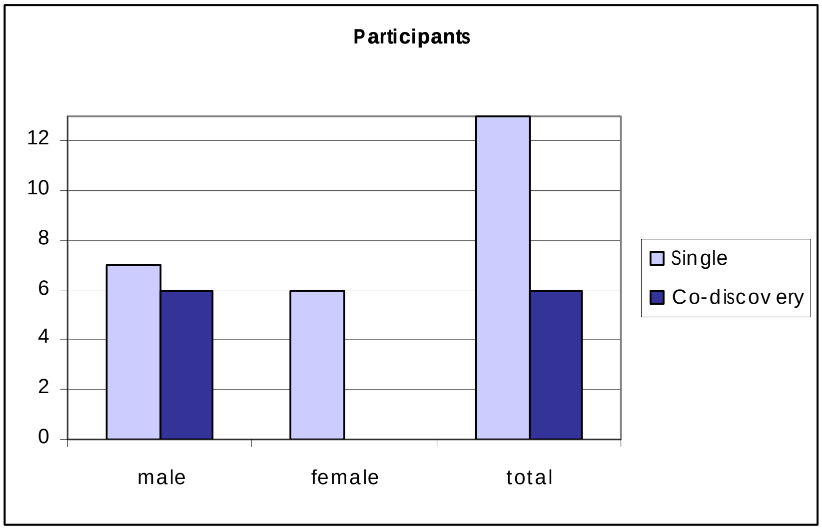


Figure 14: Chart over the number of participants and gender.

There were a total of 19 participants divided into two groups; single and co-discovery. The group with single participants contained 13 participants that did the test single-handedly. The group with co-discovery contained 6 participants that did the test in pairs.

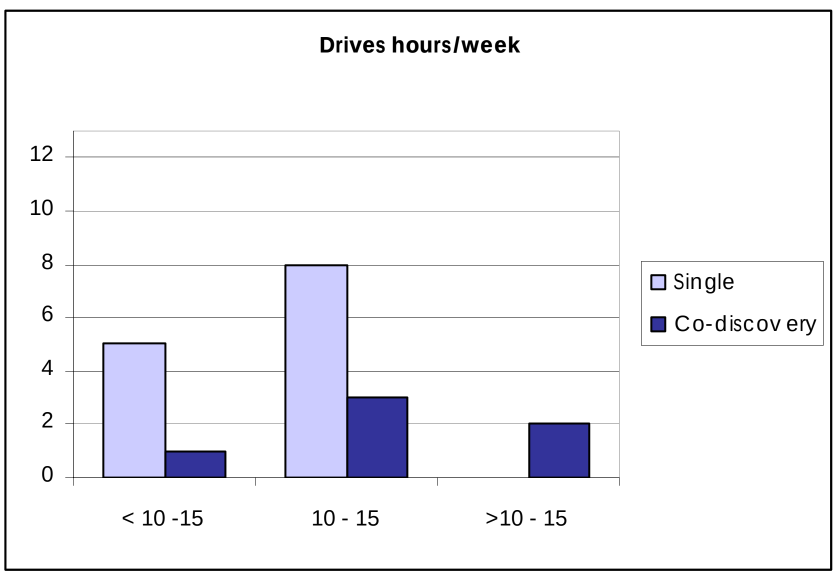


Figure 15: Chart over the participants driving habits.

The participants were all experienced drivers and drove on a weekly basis.

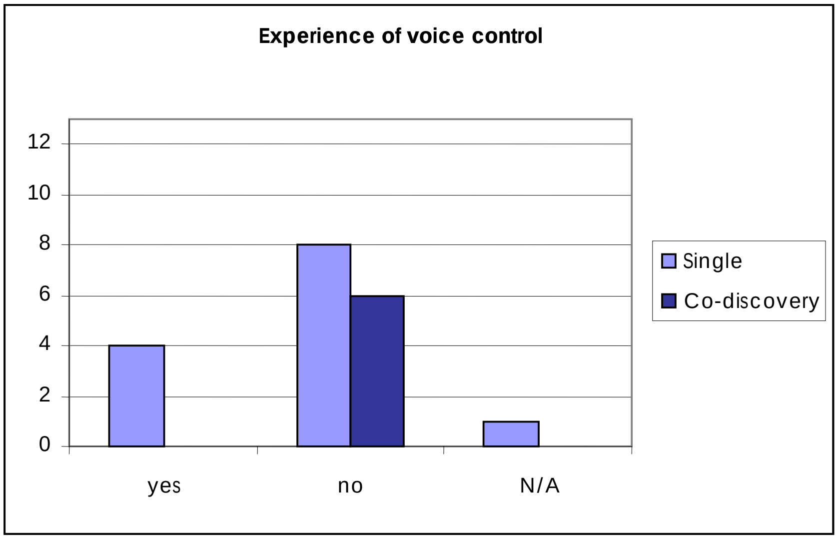


Figure 16: Chart over the participants' prior experience of voice control systems.

The participants were asked if they had used any voice controlled systems before. Few of the participants had any experience of using voice to control a system.

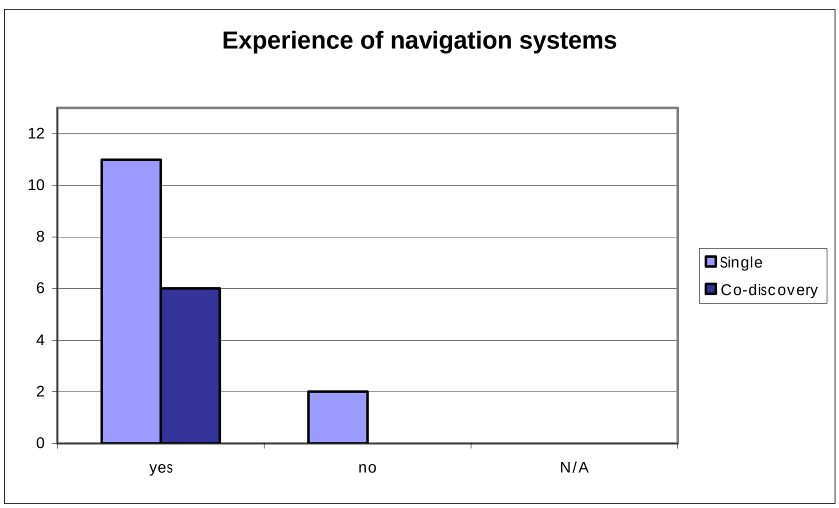


Figure 17: Chart over the participants' experience of Volvo's navigation system.

Most of the participants had tried Volvos (RTI) navigation system before. Some of them had it in their current car.

4.5.6 Pilot tests

Before the real tests took place two pilot tests (one single test and one co-discovery test) were held to see how they would work. Some minor changes of the test were made due to the results of the pilot tests. The observation protocols were changed so that they were easier to read and fill in. A redesign of some of the tasks and some of the interview questions were made to avoid misinterpretations.

4.5.7 Constitution of test

Introduction to test

The test started with a brief introduction to both the navigation system and to the voice control prototype followed by letting the participant try the system out for approximately 5-10 min. The participants were told that they could choose to perform a task either manually or with voice. They were reminded that road traffic safety took priority at all times so if a task was too demanding while driving, it should be halted. The same was applied during high workload traffic conditions.

At this point the participant was asked to start driving. Once on the road the participant, that is; the driver was asked to do the first task.

Test tasks

All tasks were presented as questions. For instance, the participants were asked how they would go about setting a destination. The test constituted of a total of 9 questions. When a driver solved a task in one modality the person was asked to perform a similar task in the other modality (manual/voice). This experience was later referred to in the interview so that the driver could make a comparison and declare a preferred modality to a task. The questions were not always asked in the same order, depending on where the vehicle was situated geographically and how the tasks were completed.

The task analysis outlined what could constitute a primary task for the voice recognition. The questions and tasks of the user test were designed to highlight these areas of the system interaction.

The co-discovery tests were conducted like the tests with one single participant but the driver and passenger switched places after task number nine was completed (bear in mind that the questions were not all asked in order).

The users were asked to perform the following tasks:

1. **From the list with the last ten destinations, set the address Temperaturgatan 2 as a destination.** To set a destination is a primary task in navigation. As today there is no possibility to set a new address as a destination with voice, the users were asked to do the most similar thing that is setting a destination from the list with the last ten destinations.

2. **Find out where you are.** The information about what street you are driving on is shown on the display. The aim of the task was to investigate the attitude towards the command *read position,* the need for operating the function with voice, and how easily the users found the proper command in the help menus.

3. **From the list with the last ten destinations, set the address Barometergatan 1 as a destination.** This question had the same aim as question number 1 but had to be performed by using the other modality.

4. **Change between one and two maps and then back again.** To change between maps was considered as a secondary task but by us found easy to operate with voice. The users' attitude was investigated.

5. **Lower the navigation volume.** The aim was to see if the users understood that there were two different types of volume, (navigation and voice control) to see if they could remember the long commands consisting of 4 words and if they could pronounce the commands.

6. **Find out how far it is to the destination.** The command *show remaining route* was considered difficult to use and to find (at the end of a list with 14 commands); the task was designed to get information on the users' attitude to the command and the functionality.

7. **Increase the navigation volume.** The aim was the same as the aim with question 5 but also to see if the users remembered the command from adjusting the volume the first time.

8. **Set the address Assar Gabrielssons väg as a return trip.** The same aim as in task 1 and 3. The aim was also to see what interaction modality the participants choose and if they remembered the commands.

9. **Delete all destinations.** This task was included to facilitate the test so that the participants didn't have to go to both Barometergatan and Temperaturgatan, only to the latter one. At Temperaturgatan the destination Barometergatan was deleted from the itinerary to avoid the system to continue the guiding for that address instead of guiding the participants back to the starting point at Assar Gabrielssons väg.

Test interview

After all the tasks were completed the driver was asked to go back to the starting point where an interview was conducted still sitting in the car. The interview was a qualitative, semi structured interview consisting of 13 questions. One of the aims of the interview was to get the drivers’ opinion of the system. Their opinion would include both what they thought of the current system and how they would like to improve it. Another aim was to capture their mental representation of the system (Appendix F).

During the co-discovery interviews both participants’ attitude were noted.

4.5.8 Avoiding possible errors

Some errors that can occur during test can be avoided by the awareness of them and by carefully planning the test with these in mind. Some of these are listed in the chart below.

Foreseeable possible errors	Measures taken to avoid possible errors	Unavoidable possible errors
Instrumental errors and the effects of interview	The interview was conducted in Swedish with Swedish speaking participants. The participants were given clear instructions. The participants were given time to ask questions.	However; the voice control prototype is not available in Swedish, but the navigation system is available in Swedish. To minimize confusion the navigation was set to the same language as the voice control; English. The participants had to address the system in English and this was not their mother tongue, they could experience problems due to this. Feedback was given in English and the participants could have difficulties interpreting this.
Loosing focus	A checklist was used before every test. All questions were read out from the same document with questions and tasks. Both test leader and test observer were active in the interview (the observer had the possibility to ask additional questions at the end of the interview). The test leader and the test observer had different roles during the test and the interview. The roles were switched after one week to avoid monotony.	The participants could loose focus and miss information that was given to them.
Over-confidence; participants having an answer to everything	The questions have been designed to meet the level of the participants' technical comprehension. The answers and the participants behaviour have been interpreted carefully.	The participants can still be misinterpreted by the test leader and/or the test observer.
Environmental differences between tests	All tests were conducted in the same vehicle. All tests were conducted along practically the same route (the participants were asked to drive towards the same destinations). All tests were conducted during daytime (daylight). No test was conducted with snow on the roads (more luck than planning). The test leader and the test observer kept to their roles during the whole test.	There were some tests conducted during bright sunlight and some during heavy rain resulting in slighter darker conditions. This could affect visibility on the navigation display. Many participants were lost and did not always follow the guiding to the preset destination. This resulted in the fact that the test environment varied.

5 RESULTS

5.1 USER ANALYSIS

To know more about the users of the voice control system the user analysis below was made. To create the user analysis customer information for the Volvo XC90 was studied and interviews were held with persons belonging to the target customer group of the car.

5.1.1 Delimitations

The voice control system is only optional for the driver and not to be used by the passengers. This excludes the passenger as a potential user.

Because being gravely physically handicapped requires a special car to meet their special needs, these users fall outside of the general category of potential Volvo XC90 users.

The main market for the Volvo XC90 is the USA without excluding Europe. This means that the users speak and understand different languages, not necessarily meaning they are bilingual.

The voice recognition that is evaluated only takes input in English therefore it is the only language considered.

Due to the fact that a driver’s license is needed we assume that the user is literate.

5.1.2 User expectations of voice control

Users might believe that they can use natural language to communicate with the system. Some users who relate to voice control as if it would operate the same way as in many Sci-Fi movies might get disappointed. Most voice control systems can only identify a number of predefined commands which compromises the user’s alternatives of addressing it. Using a PTT-button is also something that might take time to get used to as well as turn taking in the dialogue with the system. This is a result of the fact that voice control systems have to be activated by a PTT-button or by a command to recognize the user’s utterances.

Using a TTS that has prerecorded messages from a human might give the impression to the user that the system understands natural language and might respond to it with replies that the system can not handle and understand.

5.1.3 Risk Factors

There is a potential risk that if the system is poorly designed it can add stress on the user instead of simplifying the interaction. Added stress will affect the driver’s attention and traffic safety. For example designing deep dialogue structures that when interrupted can not be reentered where the interruption was made, obliges the user to start from the beginning. This may result in the fact that the user feels pressure to complete the many levels of steps in the task rather than focusing on the primary task that is maneuvering the vehicle. Another example of poor design that can affect traffic safety is when the user has to remember more than 7 commands at one time (see chapter 3.5.2).

5.1.4 Systems used today

There is a possibility that the user has experience in using voice control from other voice controlled navigation systems in other vehicles or other voice controlled applications.

There are several navigation systems on the market today that have voice control as well as there are navigation systems hat have no voice control to aid the user. Here are some brands that have navigation systems with voice control.

Brand	Manufacturer	Functionality
Jaguar X-type	Visteon	Audio, Phone, Climate Control, Navigation
Lexus LS430	Denso	Volume, Navigation
Lincoln 2003MY	Denso	Audio, Navigation
Nissan Infinity Q45, M45	Visteon	Audio, Climate Control, Cellular Phone and Navigation Systems
Honda Accords (2003 model)	IBM Embedded ViaVoice	Navigation
BMW Land Rover L322	Temic	Phone, Note Pad, Navigation

5.1.5 Target group

The car, Volvo XC90, is designed for:

- People with a driver’s license
- People who can afford the car
- Well educated people
- In the US; mostly housewives with a home to care for and children. Women around their thirties and middle aged women.
- In Sweden; male, middle age, directors
- The users probably do a lot of things during their spare time

5.1.6 Situation

- In-vehicle environment
- While driving and standing still
- Stationary in the vehicle
- On short and long trips (mostly to work or to school)
- Driving in the city or in the country-side
- In critical situations (such as under stress and time pressure)
- Driving several times a week or daily

5.1.7 Environment

- Noisy environment (In-vehicle)
- Maybe with passengers (often children) who talk and distract
- During day-time and night-time light
- When hands are occupied with other tasks, such as steering
- While talking on the mobile phone
- While parking and the park assistance system is sounding (extreme situation)

5.1.8 Level of competence

- Computer literate
- No experience with voice control systems
- As the system is an optional “add-on” we assume that the user has a positive attitude towards new technology.

5.1.9 Motivation

- It’s an optional use
- To increase safety
- To find and go to a destination/location with greater ease

5.1.10 Does the user perform the task today?

- How?
  - Not at all with voice
  - Manually

- Why this way?
  - The Voice Control is not yet available on the market.

- Why change the current way?
  - To enhance traffic safety

- What should be kept?
  - Not all tasks are suitable for voice control. Tasks that are better performed manually will remain that way.

- How can an assumption be made that the new system will change the user’s behavior?
  - The user will have more liberty to use his hands for the driving task than with manual interaction.

- Is the user motivated to change his/her behavior?
  - As a voice control system entails a heavy cost, the user will probably be motivated to change his behavior.

5.2 TASK ANALYSIS

The task analysis was made to get an image of what elements and actions are involved in navigation. It explains navigation in general and not only voice controlled navigation. The functional requirements are specified for an in-vehicle navigation aid using a display. The task analysis was among other things used for deciding on what functions to operate with voice.

5.2.1 Definition of navigation

This is how we define the act of navigating:

To navigate is how to go from A → B

where:

- A = starting point
- B = destination
- → = path

by using:

- information in the world
- known information
- information given from another person

Navigation is a planning process of how to get from one point to another. It starts before you start moving and only ends when you have reached the destination. We have chosen to divide the navigation task into two units; planning that occurs before starting the transportation that we call *pre-navigation* and planning that occurs during the transportation that we call *navigation during transportation.*

The pre-navigation

The planning process looks different when a person knows the environment and when the environment is new to the person. Certain parts of the pre-navigation are iterated during the transportation. Below is described actions taken in the pre-navigation.

- To decide or identify where A is (where you are)
- To decide or identify where B is (your destination)
- To think of how to get there, decide on →
  - road
  - transportation means
- To think of places to avoid (e.g. a tricky tunnel)
- To think of desired places to pass/stop at
- To estimate time of travel
- To create a rough idea of places that will be passed (a bank, train station or tunnel …)
- To think of an eventual return trip

Navigation during transportation

During the transportation there is an iteration of the identification of your current position (knowing where you are) to assure that you are on the right track. The level of recognition is due to whether the area is known or not. For example if you are going somewhere where you have never been before you might not have an accurate image of what to expect, like a certain color on a house on the way. But you might have a graphical image like the one from a map, maybe expecting a roundabout and therefore searching for one in the environment. During the transportation the navigation is revised as an iterative action, an alternative route might be taken when a traffic jam is detected and so on.

5.2.2 Functional demands

These are our functional demands on the system:

Definitions

Address = street address

Category = e.g. restaurant, hospital etc.

Stored location = a location already stored and named

Waypoint = a location to stop at before the destination

A route = a path with many stops

To set a starting point, waypoint and a destination

- by address
- by category
- by stored location
- by pointing

To store a location or route

To find the stored location or route

- by name
- by map (plotting the locations or routes)

To easily remove a

- starting point
- waypoint
- destination

- route
- stored location

To calculate a route

- the shortest
- the fastest
- according to different route options (avoid ferry, only bigger roads…)
- calculate time to distance

To get guidance

- visually
- auditory

Pause and continue guiding instructions

Map settings

- brightness
- 1 or 2 maps (for different levels of detail)
- zoom in/out
- heading up/north

5.2.3 Usability demands

Feedback

- The user shall get enough feedback, both system feedback and error feedback (too much feedback can annoy the user and lack of feedback may cause the user to have problems understanding the system).
- The feedback shall be informative and easy to understand.

Error handling

- Prevent errors.
- Undo-function to be able to undo the last action taken.
- Provide proper error information and information about how to handle these errors.
- Supply alternative guesses.
- Permit easy reversal of acts.
- Avoid repetitive error messages.

Consistency

- The Graphical User Interface should correspond to the Speech User Interface.

Transparency

- The user should be able to understand the functions within the system.
- The user should know where they are in the system structure.

Flexibility

- The system should be adaptive to the user's level of competence (removable hints see Lexicon).
- The user should be able to choose the information in the system.
- The user should be able to choose interaction modality.

Adaptability

- The system should be able to detect when the driver is in a critical situation and adapt the information flow.

Effectiveness

- The system should make the task easier.
- The system should provide the user with shortcuts.

Learnability

- The system should be easy to learn and provide enough information/help to handle the basic functions without using a paper manual.

Reduce short-term memory usage

- The system should not increase the mental workload.

Speak user's language

- By using familiar terminology and syntax.

5.3 RESULTS FROM THE USER TEST IN THE VOLVO XC90

When talking about a participant the pronoun "he" is used for both men and women.

5.3.1 Dialogue

The spoken interaction between the user and the system does not include the voice guiding of the navigation system.

The participants had problems knowing how to reply to questions asked by the system and at some points the system gave instructions that were difficult to understand. One participant expressed that he didn't know how to answer when asked if he wanted to be guided to a certain street:

> *"Hur ska jag konfirmera det då?"*
> *(How should I confirm that?)*

During one of the co-discovery tests the person driving asked how to proceed when he didn't understand how to answer the system. His test partner told him how to make a reply.

There were times when the participants felt the dialogue was too slow and wanted to speed it up. At times the participants failed to use the system efficiently due to the current dialogue design. Several participants asked if there was a possibility to scroll through the voice menus. They also asked how to move backwards in the voice menus:

> *"Kan man skrolla?" and "Hur går man tillbaka?"*
> *(Is it possible to scroll?) and (How do you go back?)*

When one of the participants was told he had to start listening to the menu from the beginning he expressed dissatisfaction and said:

> *"Jobbigt att man måste tillbaka"*
> *(It's tough that you have to go back)*

5.3.2 Commands

The system has a number of commands that the user can say to instruct the system. These instructions are mapped to predefined actions.

The participants had difficulties remembering some long commands and many commands were presented at one time:

> *"Svårt hålla i huvudet vad man ska säga för att komma åt olika funktioner"*
> *(It's hard to remember what to say to access the different functions)*

The way that the commands were grouped was confusing for the participants. They also had problems with some of the vocabulary, even though they spoke and understood English. Some of the problems were how to interpret the given vocabulary and others were what vocabulary that could apply as a command which is exemplified by one participant saying:

> *"…lite svårt att komma ihåg kommandona när de är på engelska"*
> *(…a bit hard to remember the commands when they are in English)*

Another problem was that the participants often tried using commands that did not exist such as "distance to destination". There were also problems understanding the commands, what functions they represented, and difficulties remembering these commands. Two participants who were trying to find the command for showing the remaining distance to the destination (task number 6, *show remaining route*) said that they didn't find any of the commands appropriate:

> *"Var var den då?"*
> *(Where could that one be?)*
>
> *"Det var ju värre...förstod aldrig att det fanns något som passade"*
> *(That was harder…I never understood there to be anything appropriate)*

Two participants said "Where am I?" when they tried to make the system read out their position. There were several alternative commands uttered by the participants for initiating the help function; "What should I say?", "What shall I say? ", and "What can I do?". These commands were recognized and properly interpreted by the system.

The system had problems recognizing certain commands that were frequently used.

5.3.3 Help Menu

The help menu can be accessed with the command What Can I Say. This help menu is where the user is introduced to the commands and the commands are grouped into three main categories.

Compared to other systems; the Volvo Voice Control Prototype does not have a menu other than the help menu. The commands are not accessed through a menu but presented in the help menu. The user is never informed of how to access the help menu.

To access the help menu the command WCIS (*What Can I Say?)* has to be used every time. The submenus in the help menu could be accessed by saying the numbers *one*, *two* or *three* (one number for each category). This is something the participants misinterpreted and they tried to access the categories by either saying the name of the category (instead of the number) or saying the number without first entering the help menu by saying *WCIS* (saying the command outside of the help menu). The latter is illustrated by the quotes:

> *“Måste man säga WCIS?” and “Finns det genväg till 2?”*
> *(Do you have to say WCIS?) and (Are there any scortcuts to 2?)*

The participants did not always bother listening through the whole list of commands, as one of the categories consists of 14 commands. One participant said:

> *“Det var väldigt många alternativ att lyssna igenom”*
> *(There were a lot of alternatives to listen to)*

The same participant later listened to the help submenu *General commands* and gave a command to the system followed by a comment that he had to say something to get the system to be quiet:

> *“Tvungen att välja något för att få tyst på den”*
> *(Had to choose something to get it to shut up)*

Some participants used the help menu as a regular menu to access commands and didn’t understand that the commands could be used as shortcuts outside the help menu.

5.3.4 Feedback

The participants were confused when the system, in certain parts, lacked feedback. This happened for example when only visual feedback was given when switching from one to two maps or the other way around. If the user had guiding instructions blocking the map view they could not get any clues to whether the map view had changed or not since no auditory feedback was given.

Some participants had problems interpreting certain feedback. For example when altering the volume and auditory feedback was given. The following was said about the feedback given when changing volume indicating that the participants had difficulties interpreting the feedback;

> *“Är den höjd eller inte, det vet ju inte jag” and “Var det konfirmation det?”*
> *(Is it louder or not, I don’t know that do I) and (Was that the confirmation?)*

> *“Går den ner ett snäpp eller?”*
> *(Does it lower the volume one level or what?)*

The way feedback was given (too quickly or too slowly) was a cause for confusion. Some participants were uncertain if the system had taken their command. Some repeated the command before the system had replied that the command was not recognized.

5.3.5 Priority

In a situation where the participants had to make a turn into another road, many of them stopped talking until they had completed the turn. If instructions were given to them by the system at the same time as they were turning the participants would ignore them. If given by the test leader they would ask the test leader to wait with the instructions until they had completed the turn. One participant indicated that he wanted to finish turning before starting with the task he was asked to do:

> *“Får jag svänga färdigt först?”*
> *(Can I complete the turn first?)*

Another participant who was listening to the help menu and had difficulties to listen and make a turn at the same time said:

> *“Man missar ju helt i och med att man svänger”*
> *(You miss it completely when you are turning)*

5.3.6 General problems

Many of the participants tried to solve parts of a voice task by using the manual buttons. This resulted in the system cancelling the action as today a task cannot be solved by combining both modalities. Another attempt for manual interaction with the voice control system was when the participants tried to navigate through the help menu using the *Back*- and *Enter*-buttons, this resulting in exiting the help menu.

The English Text-To-Speech engine was found to be poor and some of the participants laughed when they heard it trying to pronounce Swedish street names. Some of the comments about the TTS were:

> *"Juste Svenska"*
> (N*ice Swedish* (said with irony))

> *"Det är Hipp-Hipp varning på den..."*
> (*Sounds like Hipp-Hipp...* (This quote refers to a Swedish TV-show called "Hipp-Hipp" were they make fun of Americans speaking Swedish))

On the other hand the participants found the voices for the voice control and the voice guiding to be pleasant.

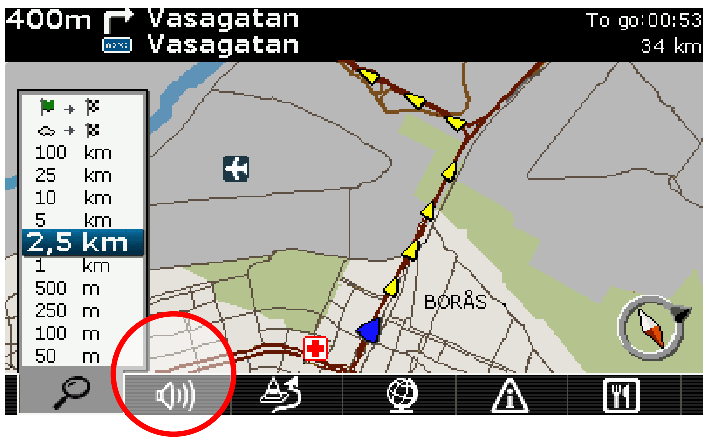


Figure 18: Shows the "quick menu" with the icon for voice guidance marked with a red circle.

The icon for voice guidance looks a lot like a loud speaker and it was mistaken for an icon for volume. This happened when the participants were asked to alter the navigation or voice control volume manually.

The voice guidance is prerecorded messages in the navigation system. The participants needed reassurance that the voice guidance was functioning properly when it didn't give directions to drive straight ahead at road crossings. Although this is part of the navigation system rather than the voice control prototype the participants did not make a difference between the guidance voice in the navigation system and the voice of the voice control prototype. This difference was not noticed even though the navigation voice had a British accent and the voice for the voice control an American accent. The following quotes demonstrate the drivers' insecurity when they expected instructions on how to drive:

> *"Ska jag köra rätt fram om den inte säger något?"*
> *(Do I drive straight ahead if it doesn't say anything?)*

> *"Har vi ingen röst? Försvann den på vägen? Sån som säger var man ska svänga och så…tryckte vi bort den?"*
> *(Don't we have any voice (guidance)? Did it disappear along the way? The kind that tells you were to turn and so on…did we put it off?)*

There were a considerable number of participants who tried to set a destination by saying street names as commands;" I want to go to Temperaturgatan 2", "Barometergatan" and so on.

When the centre stack display shows the text *Listening* the user does not need to press the PTT-button again before saying a command. The text indicates that the system is still actively listening for a command and has not yet timed out. Some times the centre stack display showed the text *Listening* for a very long period without actively listening for a command; the system had in reality timed out. This is a bug in the system which caused some confusion among some of the participants.

A few participants seemed to have great difficulties with what we call "timing " (the term also appears in some literature on human spoken dialogues and interaction management (Taylor, Néel, Bouwhuis, 2000)) which means that they said the commands either too soon or too late for the recognition engine to capture the whole command. The majority of all participants managed to interact with the PTT-button smoothly.

Some of the participants expressed that they experienced it as dangerous to look frequently at the display. One participant said he had to concentrate in order to avoid having an accident while manually trying to change between maps:

> *"Nu gäller det att inte köra på något"*
> *(I'd better not hit anything)*

Some participants expressed that they focused too much on the display. One person drove very slowly while accomplishing the task (task 1) manually and accelerated when finished.

5.3.7 Unexpected incidents

The average participant failed to follow the guiding instruction approximately about three times during the test. One person said (translated to English): "There are too many things happening at a time" when he failed to follow the driving instructions given by the navigation system. The same person later stepped heavily on the brakes in front of a pedestrian crossing.

One person didn’t notice that the traffic light turned green during interaction with the system but was reminded when a driver behind him started to honk.

During one of the co-discovery tests the driver stepped heavily on the brakes at a stop light when it turned red. The driver was trying to understand why the system didn’t want to accept his attempts to confirm the guiding request by saying “yes”.

One person drove up to the verge and stopped, talking a lot about the system. When later asked why he had stopped he answered that he didn’t know what he was supposed to do in the test. This participant seemed reluctant to wait for instructions from the test leader and preferred testing the system in his own manner. Prior to driving up to the verge and stopping he seemed to have problems managing the situation and seemed confused.

Some participants wanted to stop the car during the test and were told that this was okay.

One participant bumped into a traffic island when searching the graphical menus for a setting option and at the same time making a turn. The same person later wanted to stop the car and continue the test with the car standing still. Another participant drove close to a traffic island while explaining to the test leader that he wanted the map to show heading-up-mode.

The recognition rate for one of the participants was very low. The participant tried to enter menu number two in the help function by saying the command “two” ten times in a row without being understood by the system. Due to the problems with recognition, difficulties in following the guiding instructions, and to spare the test person from a frustrating and unpleasant situation, the test was interrupted.

5.4 A SUMMARY OF THE INTERVIEWS

5.4.1 Mental model

Half of the participants in the single test considered the voice control prototype and the navigation system as one system. The other half of the single test group and the entire co-discovery group saw the two systems as two separate systems.

5.4.2 Modality

Voice interaction versus manual interaction

When the participants in the single test were asked about what interaction modality they preferred there was an equal distribution between the two modalities.

Two participants said that they could interact in either modality, without specific preferences, depending on the task. One of the reasons for choosing voice control was that manual interaction involved more fiddling. With voice, only one button is used (the PTT-button) instead of three buttons. However one participant remarked that the PTT-button is difficult to use when turning the steering wheel.

Some participants said they preferred manual interaction as they were used to this modality. Others stressed the difficulty to remember what commands to use with voice.

Participants in both test groups in general had a positive attitude towards voice as a modality, saying they were willing to learn how to use voice control and that they would probably use it if they had it in their own car.

Tasks considered to be best performed with voice control were tasks including setting a destination or changing the volume. It also showed that it is important that the commands are easy to remember. Participants considered the voice control to be smooth when they could remember and use the command directly. The greatest disadvantages with voice control were considered to be the long menus in the help function and commands consisting of more than two words.

Manual interaction was preferred for different types of settings such as changing map scale, altering between one and two maps, altering map colours and so on. If a task only demands pushing a button once or twice the task is best performed manually.

Tasks that the participants in the co-discovery test felt were better solved with vocal interaction were tasks solved by using simple commands such as switching between one and two maps (*one map* and *two maps*) and tasks that manually require many steps to solve.

Being able to switch between modalities was something positive as they claimed that voice control was a good complement to the navigation system. One modality would not necessarily rule out the other.

Multimodality within tasks

Some navigation systems have a touch screen, this navigation system does not; it has manual controls placed behind the steering wheel; the Back-, Enter- buttons and the four-way switch. The voice control gives visual feedback on the display such as hit lists. The user can not use the manual buttons to scroll or choose alternatives in any lists given from the voice control system.

When the **participants in the co-discovery group** were asked if they would like to be able to solve a task both manually and with voice, such as setting a new destination, they replied that they would. One participant wanted to be able to use the *Back*- and *Enter*-buttons, used to navigate through the navigation system, to navigate in the voice control help menu (*WCIS*). Most of the participants, when asked, agreed that they would like to be able to change the volume by using the voice command *volume* to activate the function and then manually setting the right level with the four-way switch.

Multimodality within a task was considered applicable by the **participants in the singles group** when making a selection from a list. The list could be retrieved by using voice and then a selection could be made by using a button. For example, this way of interaction could be used when setting a destination. A majority of the participants agreed that it would be a good idea to specify which volume (voice control volume or navigation volume) with a voice command and then use the four-way switch to adjust it. That way it would be possible to change the volume several steps at a time without being obliged to say the entire command all over.

5.4.3 Feedback in the centre stack display

More than half of the group consisting of single test participants and all of the co-discovery participants had noticed the information on the status of the voice control prototype (Push -To-Talk, Listening, Processing and Cancelled) shown in the centre stack display. Both those who had and those who hadn’t noticed the information disliked that the information was situated in the centre stack display. This was motivated by the fact that the attention is drawn to two different places and that the centre stack display is placed too low. The information would preferably be presented higher up and closer to the field of vision when focusing the eyes on the road. The navigation display was considered to be the best place for showing this information but some participants argued that the navigation display already contained enough information. Instead of using a text a diode and the colours red and green could indicate the status of the system. Other suggestions were to show the status information on the Driver Information Module (DIM)-display (Figure 19), close to the speedometer or on the windscreen.

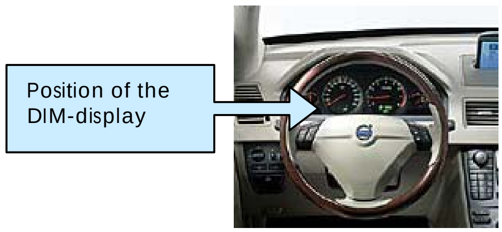


Figure 19 : Shows the position of the DIM-display in the Volvo XC90.

5.4.4 Talking to the system

The participants in both test groups had unanimously answered yes to the question if they had understood how to use the PTT-button, even though some of them had difficulties interacting with the voice control prototype. Almost all participants knew that it's possible to barge in during a played prompt and how to do it. All participants in the co-discovery group also barged in during the test.

They were all satisfied with how well their commands were recognized but they realized that they were required to articulate and not mumble in order to be certain that the speech recognition engine would understand their commands. Several participants mentioned that the system didn't understand the command "two" (this command can only be used within the help function). One participant blamed himself for not speaking clearly enough, another participant said he made an effort to speak clearly.

5.4.5 Commands

Most of the participants thought the commands were too long. Especially the commands for changing volumes, which contain four words, were considered too long by almost everyone. Some of the participants wished the commands to contain a maximum of two words.

The commands were not considered difficult to pronounce.

One participant preferred saying volume "plus/minus", another person preferred saying "higher/lower", instead of "louder/softer".

One participant suggested asking the system "Where am I?" instead of using the command *read position*.

One participant didn't find the commands *one, two,* and *three* informative. These commands are used to enter the three submenus in the help menu. The commands don't indicate the content in the menus.

The command *WCIS* was considered intuitive and easy to remember by some users. Others asked for the possibility to say "help" to start the help function.

5.4.6 Dialogue

Two thirds of the participants in the single group and all of the co-discovery participants agreed that more dialogue in the interaction (like in the Jaguar AVP) would be a good thing. They thought that it could help take some of the mental workload off the user since the system could help a user who doesn't remember the commands. They would like the possibility to ask questions to the system and the system to identify what they want to do and give instructions on how they can do it. One participant wanted the system to use dialogue for error handling, e.g. if the system is unsure of what the user wants to do it can ask "Do you mean this or that*?"*.

The participants emphasized that more dialogue should only be added with plenty of care as they feared that making the system more dialogue based could make it seem too "chatty".

Those who didn't want a more wizard like dialogue believed that it is more efficient to use commands.

5.4.7 Help Function

A majority of the participants claimed that they found the information in the help menu where they expected it to be found. Two participants brought up that the menu name *General commands* is a very general description of the menu's content.
All participants except one considered the submenus to be too long, in particular submenu number two; *Guiding and Destination* which contains 14 commands. This caused a lot of confusion for the participants as they said that after having heard some of the commands they forgot what command that they were looking for to begin with.

One suggestion is to have one help function for novice users and another for frequent users. Another proposal is to remove unnecessary commands from the menus.

One participant emphasized the lack of match between the content and the structure in the help menu for the voice control and the navigation system's graphical interface.

Another participant wanted the GUI and the SUI to be as similar as possible.

When participants in the single test were asked the question “Would you like the help menus for the voice control to be shown graphically as well?” most of the users answered yes and argued that it would ease the perception of the voice menus. Two participants claimed that it would be dangerous to show the menus on the navigation display. For the co-discovery group no conclusive answer could be given.

A context aware help function was proposed, it would only inform the user on available commands and not the entire set of commands within the system.

5.4.8 Positive aspects on voice control

Shortcuts were judged to be a positive aspect of voice control. Some participants thought that using voice control to operate a navigation system could make driving safer. One reason for this was that it is easier than using buttons and that the driver can keep his eyes on the road. Voice control was claimed to be more natural than manual interaction.

Thoughts articulated about voice control were (translated to English); “*cool*”, “*quite good*”, “*I like the system*”, “*it is a funny novelty*”.

5.4.9 Negative aspects on voice control

Many participants made complaints about the TTS, they found it hard to understand and the voice unpleasant. Some participants felt insecure due to the lack of feedback and found the long menus frustrating.
Some thought that the commands were hard to remember and one participant thought he would probably not learn more than 20 % of the commands in the system.

5.4.10 Pause function

A majority of the participants wanted to be able to make a pause in an ongoing voice interaction. It was considered to be useful if under stress, when making a turn, when talking to a passenger or when there is a phone call.

The co-discovery group preferred pressing a button but a narrow majority of the single test participants wanted to use only a word for pausing the system. Words suggested for pausing the system were; *pause*, *wait*, *stop* and *hold*. To restart the interaction the words *continue* and *proceed* were recommended by the participants.

Only one of the participants disagreed on the advantage of pausing, meaning that being able to repeat the prompt is better (something that can be done in the current version of the voice control but that the user for whatever reason missed during the test). The participant also said that even better than being able to repeat a spoken prompt is to have a system that is aware of the user’s current traffic situation and for instance, avoid playing an informative prompt while the driver is turning.

5.4.11 The participants suggestions for improvement

The participants would preferably speak to the system in Swedish (mother tongue). They also emphasized the importance of keeping the commands short.

Some participants wanted the voice control help menu additionally to be shown visually on the navigation display.

The participants wanted the possibility to set a new destination by saying the street name or by spelling it. One participant said he wanted the system to understand whatever he said, speaking in whatever language.

Improvements could also be made by being able to zoom many levels with a voice command; currently this can only be done manually as the voice commands only zooms one step at a time. Some thought that the system should be redesigned and utilize *forcing function* such as not letting the user start doing a task if the user is about to make a turn. Some wanted that customers waiting for their ordered car to be delivered should be offered an introduction-CD for voice control so they could get a head start on the system functionality.

5.5 HOW WELL DOES THE JAGUAR AVP MEET THE USABILITY DEMANDS?

The voice control prototype for Jaguar, called Active Voice Portal, is somewhat different from the voice control prototype in Volvo's XC90. In the XC90 only the navigation system can be controlled by voice at this point, where as Jaguar AVP has voice control for audio, climate control, communication and so on. The main difference lies in that the voice control for Jaguar is not yet implemented in a vehicle while the voice control prototype in XC90 is. This made the test conducted on the Jaguar AVP different from the test conducted on the Volvo Voice Control Prototype. To see how the test was conducted please see chapter 4.4.

Some of the problems in evaluating Jaguar's voice control lie in the fact that we are not always certain if a weakness is due to system failure, that the system is at its early prototype stages (all functions are not yet implemented) or that there is a lack in the dialogue design (SUI).

5.5.1 Dialogue

Jaguar's AVP is a dialogue based system. This means that the user doesn't need to remember a vast amount of commands; these are given to the user during the dialogue. But if the user does remember commands these commands can be used as shortcuts during the interaction. Once a dialogue is started the user is asked a number of questions in order to specify what the user wants to do. This means that the system is the one taking the initiative in the dialogue and not the user.

The listening of the voice recognition engine is started by pressing a button. However, the button does not have to be pressed each time the user wants to give the system a command. The user only needs to press the button to start the dialogue, the system then listens for the duration of the dialogue. One of the advantages is that the user is not dependent on how the steering wheel is currently positioned. A major disadvantage is that any noise can be interpreted as a potential command.

The system has a natural flow in its dialogue and uses a familiar vocabulary and a comprehensible grammar. The language that is used gives an impression of the system to be very polite and the system has a very pleasant voice (this is a prerecorded human voice).

5.5.2 Menu structure

The Jaguar AVP has a main menu. The user presses a button to start the system and gets into the main menu. The main menu does not have to be requested and its' contents is read out when the AVP is started. This means that the system is ready to use, a manual or other instruction on how to interact with the system is not necessary.

In the *more options* menu the list starts with "This list of commands is long…". The users' attitude towards the system can be affected negatively by such utterances. It's a poor solution to warn the user of a problem instead of improving the design; in this case a redesign of the menu structure is required to make the command list shorter and less demanding for the short-term memory.

The help menu in the Jaguar AVP is context aware and describes the different functions in more details.

5.5.3 Effectiveness

Shortcuts can be used throughout the interaction and are intuitively used at the short pauses provided by the system in a dialogue before the system gives instructions on how to proceed.

Unfortunately the system does not always provide the commands that can be used as shortcuts which leave the user with endless tryouts.

Due to the fact that there at this moment is no Graphical User Interface adapted to interact with the SUI the system might not supply the user with total confidence and at times leaves the user dependent on time consuming dialogues. In a GUI supported SUI system the information is presented in two modalities; visual and auditory. Sometimes the visual information can be accessed faster than the auditory information.

5.5.4 Consistency

The Jaguar AVP has problems in consistency with commands, both in how commands are presented to the user and how the user is supposed to command the system.

Some feedback messages in the system lacks in consistency which can confuse the user.
Other inconsistencies can be found in the interaction when the system asks for confirmation to a choice made by the user. Some commands work differently if used while a prompt is being played or not; for instance the command *cancel*.

The system can also instruct the user to use different commands to the same function in different parts of the dialogue. For example; the user is told in one part of the dialog that the main menu can be accessed by saying *main menu,* in another part of the dialogue the command *Jaguar* is to be used to access the main menu. Also the commands that the user can use as shortcuts do not always correspond to what the user used where a command during the dialogue, even though the function is the same.

The command *repeat* repeats a spoken prompt from the system, unfortunately not always at the same level as the user is situated in the dialogue structure but in one level above the current location of the dialogue.

5.5.5 Feedback

The tested system had no visual feedback. At some points the system also lacks informative feedback. Jaguar has a separate navigation system with a display but none was used during testing.

The long pauses in the system can be confusing for the user as it’s not always clear if the system is still listening.

5.5.6 Flexibility

The user is not able to choose the information that the system presents, menus are fixed, and interaction modality can not be changed as there is no manual interaction. The system lacks removable hints (see Lexicon).

To set a street name in a bigger city like London the user first has to tell in what suburb the street is situated or the post code. This is a weakness in the system as the user then has to know information that the navigation system should provide for the user. A visiting driver who is not familiar with the post codes of a city should be able to use the system without problems. Maybe this problem could be solved by allowing hit lists.

5.5.7 Adaptability

The system does not detect when the driver is in a critical situation, traffic wise. It has no knowledge of the driver’s traffic situation or the amount of information that the driver is handling at any given moment.

The system does detect, to a certain degree, when the user is having trouble getting recognized and gives comforting feedback or encourages the user to try again or try in a different way. This is done according to the number of times the system has problems recognizing a command.

5.5.8 Error handling and error prevention

The system does not permit easy reversal of acts. The system does not provide hit lists (neither visually, due to the fact that it offers no display, nor spoken) which is necessary to supply alternative guesses.

The system does avoid repetitive error messages and keeps track if a user is having problems with recognition. If this happens the system suggests alternative ways for the user to achieving the task.

When setting a destination, the address is set piece by piece; first the user says a town, then a suburb and finally a street name. If the user had to say all at a time there would probably be more difficulties with the recognition.

6 ANALYSIS AND IMPROVEMENT PLAN

In this chapter the results from the user test are discussed and proposals for how to improve the voice control prototype are presented. The proposals are based on the results from the user test and on the writers’ reviews of the system.

6.1 MULTIMODALITY

The users are able to choose if they want to solve a task manually or with voice control but not both at a time. Using more than one modality to solve a task might speed up the interaction and minimize the mental workload on the user. Some examples where multimodality could be applicable are:

To be able to use two modalities to solve a task. During the usability test the participants were asked if they would like to interact with the system both manually and with voice to solve a task. As an example they were presented with a way to alter the volume; they could use the voice command *volume* to activate the function and then choose the level by pressing the four-way switch. This permits the user to keep the eyes on the road. At the same time choosing manually the desired level would come more natural than saying a volume level. A majority of the participant thought this was a good idea.

To be able to choose manually from a hit list. The user can start interacting with the system by voice control and when presented with a graphical hit list, the user can choose to select an option manually and still be able to go back to interacting with the system by voice commands.

To be able to select a specific value from a list. Today, when the user wants to zoom in on a map the user can manually choose to zoom in from scale of 100 km to 50 m by pressing on the buttons four times. Changing the scale of the map with voice the user has to use the command *zoom in* but this command only zooms one step at a time, so zooming in from 100 km to 50 m would require the user to repeat the command nine times (see Figure 20).

By letting the user specify to what scale he/she would like to zoom in to, this could be avoided, for example *zoom in 50*” (this value does not need to be specified with meters as there is only one value of 50 in the list of options).

Reading out values when scrolling through a list of options. To avoid that the user has to look at the display to see a list of options the values could be read out to him when he scrolls through them manually.

This should only be done if there is no delay when reading out the options. For example the list with scale options could be read out, see the picture below:

Figure 20: Shows the zoom options in the quick menu.

To set a new destination with voice control. Currently the system does not allow the user to set a new destination with voice. Setting a new destination is something that is considered to be one of the primary tasks in the task analysis but something that is technically very hard to do with voice. Setting a new destination manually can be done by selecting country, town, street name and street number from a list of options. The options are narrowed by top-down method (see Lexicon).

One of the reasons why using voice to set a new destination is technically difficult is because of the number of commands the system has to recognize such as all the street names in a system. Just allowing the user to name any street to set a destination would implicate that all street names have to be set as possible commands. This would require a very large hard drive for storing commands and much more processing power which could be very expensive. But if the options were narrowed as they are in the manual interface and the user was allowed to interact in both modalities (manually and voice) it might be possible. This is how it could be done;

The user initiates the navigation system’s set-a-destination-function by using the voice command *set destination*. As the user usually sets destinations within the country boundaries this is by default already chosen but with the option of changing it available. To set the street name the user can choose to use either to set it manually or to set it with voice by spelling out the three first letters of the street and hence the system would present the user with a hit list. Hit lists are suitable if the user isn’t certain about the name of the street or of how to spell it. It should be optional to set a street number but this could also be set by voice. If the system also has a function that reads out options from a list the user won’t need to look at the display to set a new destination.

6.2 CONSISTENCY BETWEEN THE SUI AND THE GUI

The navigation system is in production and the voice control prototype is not, hence some users will be familiar to the structure of the navigation system.

During the test the participants were asked if they perceived the navigation system and voice control prototype as one or two systems. Half the group of participants saw the system as two systems. This means that there is a possibility that the user has two mental representations of the menu structure; one for voice control and one for the navigation system. If this is the case the user would need to use both representations to solve a task which implies a greater mental workload than using one (see mental workload in Lexicon).

The user needs to feel confident with the system. Some of the users might know how a task is solved manually in the navigation system and will probably try commands out of the vocabulary that is shown graphically. Having a SUI that resembles the GUI will work as a support for the memory.

Sound is transient and so is the speech help menu in the SUI, therefore a user needs to remember what has just been said or the information is lost. This can be compared to a GUI where the information is available to the user as long as it is presented on the display; the user only needs to glance at the display to re-access the information. It is therefore quite probable that the user will glance at the display for hints if the user forgets or misses played information from the SUI. If the GUI resembles the SUI the user can use the information presented in the GUI to recall information in the SUI.

6.3 MENU STRUCTURE

The menu should be transparent. One of the main differences between the Jaguar AVP and the Volvo Voice Control Prototype is the menu structure. The Volvo Voice Control Prototype does not have a main menu; it has a help menu that can be called upon with the command *What Can I Say*. Due to lack of feedback, many of the participants in the test used the help menu as if it was a main menu they had to enter to use the commands. The system did not inform the users that they could use the commands presented in the help menu directly, without entering the menu. As a result all these participants grew impatient with the system as they found the long menu to be tedious and time consuming.

There are two ways of improving the current menu in the Volvo Voice Control;

1. To keep the existing menu (a help menu) but restructure it by
    - Regrouping the commands into other categories.
    - Using shorter menu names that better indicate their content.
    - Only keeping commands that are essential. Some commands do not fulfil their purpose (to aid the user) and should be excluded.

2. To have a main menu that presents the main commands and a separate help menu that can help the user with more information on a specific part of the interaction. The help menu should be specific to where the user is in the menu (context aware) while the main menu should be static.

Human paced systems and introductions. The Jaguar AVP introduces the user to the computer by the main menu, hence letting the user know that he can speak to the computer. This is something that isn’t done in the Volvo Voice Control prototype. The user has to take the first step and activate the voice control through the help menu (*WCIS).* Since the Volvo Voice Control Prototype is a system that lets the user take command on when the interaction should take place (human paced) the user has to know the commands beforehand or use the help menu to figure out what commands to use. It could be a good idea to introduce the voice control to the user maybe by playing an introduction prompt. This could be done while still keeping the system human paced.

Modelling (see the guidelines in Appendix A) should be used in the introduction. The introduction should not make the user think that the system can recognize continuous speech. If the introduction is brief and terse the user will speak in the same manner.

Regrouping the commands into other categories. The average human can keep 7 ± 2 *meaningful* units of information in the short-term memory at the same time (see 3.5.2). Grouping commands with similar function together make them easier to remember. The most common commands should be placed in the beginning of the menu.

Reasonable submenus. A very large part of the participants had problems directly caused by the long submenu containing a total of 14 commands (Figure 7). Speech interfaces tax more on the user’s memory than graphic interfaces and there is no visual interface attached to the speech interface. This means that the users are able to remember about three or four menu or prompt options at a time (Weinschenk, 2000). This is far less that the system requires.

Only use necessary commands. The submenu containing 14 commands can probably be both regrouped and some commands could be excluded all together. The participants in the test who had used the navigation system before said during the interview that they rarely used all the functions in the navigation system. Voice control should be used to control only the very frequently used functions in the navigation system, the functions that are essential to the user. To see some examples of commands that are discussed to be redundant see the chapter Analysis of commands 6.5.

Having a main menu and a help menu. The Jaguar AVP has a main menu and a help menu. The commands are accessible through the menu and some commands are shortcuts and enable the user to get to a function or part of a menu directly (unfortunately the shortcuts are not presented to the user, see chapter 5.5.3. The user can call upon the help menu to get a more detailed explanation on system functionality when in doubt. The Jaguar AVP is not human paced but the developers of the Volvo Voice Control Prototype would like to keep their voice control system human paced. So how can a menu be created so that it is still human paced? One way is to let the user call for the main menu by using a command *main menu* where the commands are grouped and presented to the user. All commands that are presented should work directly, as shortcuts. The user should not need to enter the main menu to use a command. This will make the system flexible both for novice and for advanced users. To get more information on how to use voice control the user can call for the help menu by simply using the command *help.* The help menu should be context aware so that it helps the user with the current task.

6.4 RECOGNITION OF COMMANDS

Over all the speech recognition in the XC90 is satisfying. Despite the noisy environment the system's level of recognition is high. However there were some problems with the recognition of the commands uttered by the users. The user tests showed that there were problems recognizing some of the most frequent commands (*2* and *cancel*). Problems with recognition can probably be reduced if the recognition engine is further trained.

If the system can be addressed in Swedish perhaps there will be fewer problems with recognition as the users will be surer of how to pronounce the commands. This obliges the recognition engine for Swedish to be trained as well.

During the co-discovery tests the level of recognition was sometimes poorer than during the single tests as the two test persons in the co-discovery test were talking to each other when the system was listening. As the system is implemented today the passengers must think about being quiet when the recognition engine is listening for commands. To overcome this limitation another microphone is required that can better sort out sounds and utterances not coming from the driver.

6.5 ANALYSIS OF COMMANDS

Shorter commands. The commands *voice control volume softer/louder* and *navigation volume softer/louder* are too long and very hard for the users to remember.

Commands with similar functionality. On the display the information about the remaining time and distance to the destination is grouped. With voice control there is the possibility to make the navigation system read out loud the remaining time by saying *read remaining time*. The users then believe that they can say *read remaining distance* to make the system read out loud this information. That command does not exist, however there is the command *show remaining route*, which shows the remaining route on a map on the display and the remaining distance in kilometres. Today it is not possible to ask the system to read out the remaining distance. As the information on remaining distance and remaining time is similar and also grouped visually it would be reasonable to use the same way to request the information and to present it. If in a future system there will be the command *read remaining time* the system should also preferably have the command *read remaining distance*.

Avoid confusion between commands. In the user test one task was to set a destination from the list with the ten last destinations. When listening to the help menu, many users mixed up the command *stored location* with the command *set destination.* "To set a stored location, one of the last ten destinations or a return trip please say set destination. To store a location say, store location." One explanation could be that the users knew that the destination was already stored in a list so when they heard the command *stored location* they triggered on the word stored and thought that this command would open the list with the last ten destinations. To enter that list when being in the help menu the user should say *set destination* and then *last ten destinations*. Another explanation to the mix up could be that the users repeatedly heard stored/store location, which made them say it. Perhaps the Swedish users don't make a difference between a location and a destination.

The command *set destination* was mixed up with the command *directions.* The users could for example say set direction instead of set destination. Probably this confusion could be avoided if the voice control is available in Swedish. The confusion between destination and direction is perhaps due to the similarity between the words destination and direction. Both words start with the letter d and finish with –tion. The words are similar in length even though the word destination has four syllables and the word direction three syllables.

Asking questions in addition to commands. To make the system read out the street you are driving on there is the command *read position*. Both earlier user tests and the user test performed for this thesis have shown that some users wanted to say *Where am I* to get this information. This is similar to the command *What Can I Say* that is used to initiate the help function.

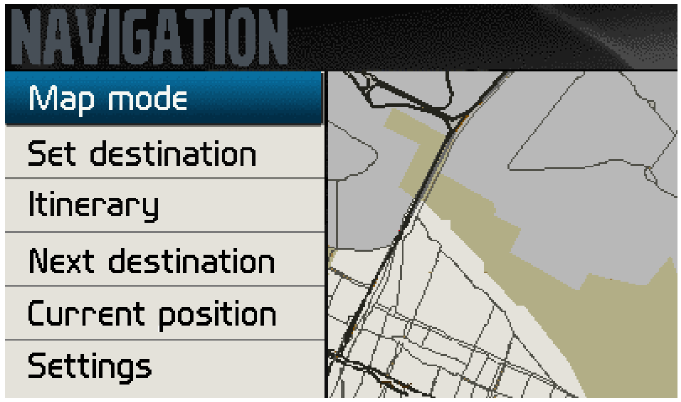


Figure 21: A view of the menu in Volvo's navigation system.

Easy map access. In the graphical menu today the option Map mode is used to display the map. This is a function that will be removed as the user should be able to go back to map mode by pressing the *Back*-button until the map mode is entered. To easily access the map without having to press the *Back*-button several times, e.g. when in the other parts of the graphical menu, it could be suitable to have a voice command; *show map*.

Today the voice command *map mode* is divergent to the command *route outline mode,* which shows a list of the next guiding instructions. When a user says *map mode* the list is removed and instead the map is displayed. However, if something else is blocking the map, such as a hit list or the graphical menu, the command *map mode* cannot be used to display the map. The command could be kept connected to the current function, but it should be possible to enter from any part of the system, showing the map on the display.

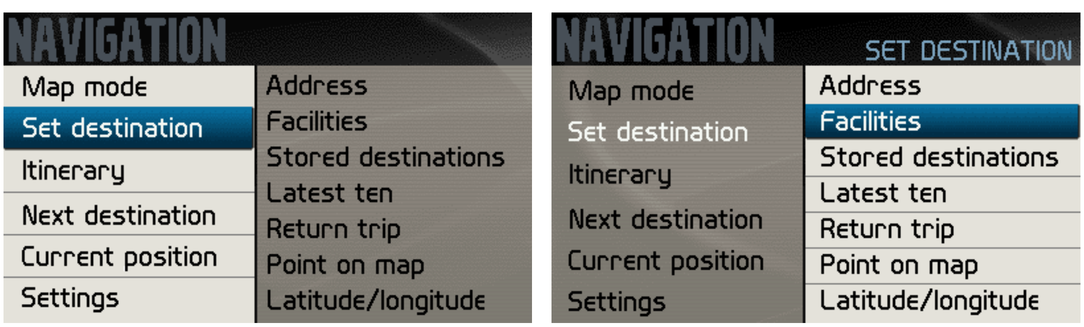


Figure 22: This picture shows the graphical view when the user selects *Set destination* (left), enters the submenu and *facilities* is marked (right).

Reading what is on the display. One way of keeping the user from looking at the display too much and being distracted is to let the user be able to have the text read out aloud. This could be useful when for example the user wants to choose a map scale or just navigate the graphical menus without having to look at the display. The command proposed is *read out* and it would read out the marked text in a menu or hit list (Figure 22). The command will start by reading out the first option in the chosen menu/hit list. If the user moves up or down with the four-way switch one step in the menu/hit list the next option will be read out until the user exits the menu/hit list.

Reading traffic information aloud. The navigation system receives traffic information that is shown graphically as text and icons that can be difficult to interpret. To attain the information the user has to go through many steps in the GUI and this requires him to focus his attention on the display. To enhance safety these traffic messages could be processed by the TTS and then read out to the user so that the driver can keep his eyes on the road. A suitable command to request this operation could be *read traffic info.*

Revising the functionality of the commands. It should be reconsidered what functions are suitable to be operated by voice control. It is not possible to set a new destination with voice. However this is something we consider to be a primary task and a highly desirable function.

It can be discussed whether the command *read remaining time* is necessary as the time to the destination always is shown on the display and easily accessed by a glance. Another command which can be discussed is *read heading* which reads out in what direction the car is travelling (north, south, east or west). This information can persistently be seen on the map and is useful when driving in the US where it is important to know in what direction one is travelling. Road signs in the USA show what road one is travelling on and in what direction e.g. Route 66 East. This information is not as valuable in Sweden and most of Europe as road signs don’t display direction by cardinal points e.g. E4 Stockholm.

6.6 COMMENTS ON UNEXPECTED INCIDENTS DURING TEST

Some of the unexpected incidents indicated that the drivers were stressed and that their attention was drawn to other things than the road. No significant difference in distraction was noticed between manual and voice interaction but this was on the other hand not the purpose of the test. It is also to be considered that most of the test participants had tried the manual interaction for Volvo’s navigation system and that some of them had the same navigation system in their own cars and hence they were used to manual interaction.

6.7 PROPORSED IMPROVEMENT PLAN

6.7.1 Redesigning the menu structure

The majority of the participants used the help menu (*WCIS)* as a main menu, not always realizing that they could use the commands directly. But also, the structure of the help menu (*WCIS*) made it hard for the participants to remember the commands. The help menu (*WCIS*) is context aware. Our recommendation is to have a main menu and a help menu. By letting the main menu be static throughout the system the user can be helped to remember its content. But a help menu is more manageable if it is context aware, as it should only give advice to the user's current problems. A system that knows where the user is and can give accurate help gives the user reassurance that his problem is solvable. We propose that the help menu should be accessed with the command: *help*. The command *What Can I Say* that previously was used to access the help function should now be used to access the main menu as it tells the user the possible commands.

The current menu structure was redesigned by regrouping the commands. Every command was evaluated; did it serve a purpose of making interaction safer or was the manual interaction faster/easier? No more than four commands are to be presented at one time. The menu should have a deeper structure rather than a shallow (see chapter 3.4.2) to minimize the user's mental workload but not too deep as that would make interaction time-consuming and the user frustrated. The results from our test were considered as well as the results from earlier tests on the voice control. This is the result;

The menu can be called upon by the commands *menu*, *main menu* or *What Can I Say*.

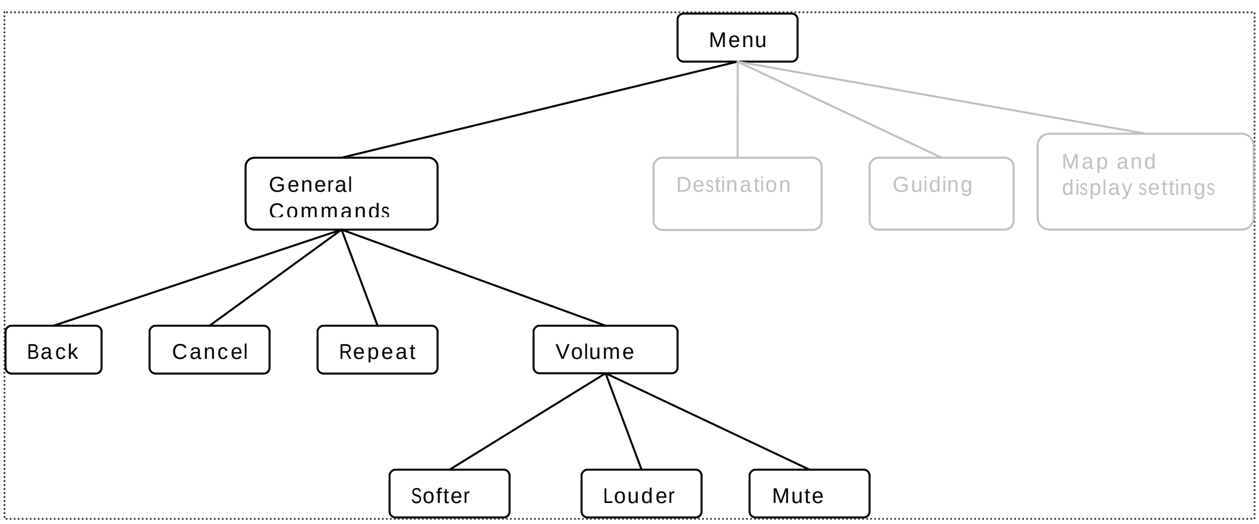


Figure 23: Proposed menu structure, submenu *General commands*.

General Commands. The commands: c*ancel and repeat* were implemented in the old system but not presented to the user. The command b*ack* is new and is used to go up a level in the dialogue just as the *Back*-button manually takes the user back one step in the interaction. It should be implemented only if the system, by training, can handle not to mistake *back* with *map*. The command *back* was recommended during the previous test 2 (see chapter 3.4.2) but it was also evident in our usability test that the participants tried to use the *Back*-button on the steering wheel to navigate through the voice menu.
The command *cancel* should be trained in the system so that the system recognizes it better than it did during the test.

The functionality of the volume command is changed. The new command alters both the navigation volume and the voice control volume independently. This function was by a test participant compared to the interior panel light, where all the panel light is dimmed with one switch. If a user finds the volume to be too loud the user will most often find all volumes (navigation and voice control) to be too loud and want to turn all of it down. The ability to alter the volumes separately should still be available to the user but not necessarily with a voice command, this could be done manually. If the navigation volume is set lower than the voice control volume manually when the car is parked, the user lowers both volumes one step if he uses the command *volume softer*. This means that the navigation volume will still be lower than the voice control volume. The early prototype had the commands *lower/higher* to set the level of volume but these commands were changed to *softer/louder* due to recognition problems. The system can however be trained to recognize the commands *higher* and *lower* but it was not done in the early prototype due to lack of time.

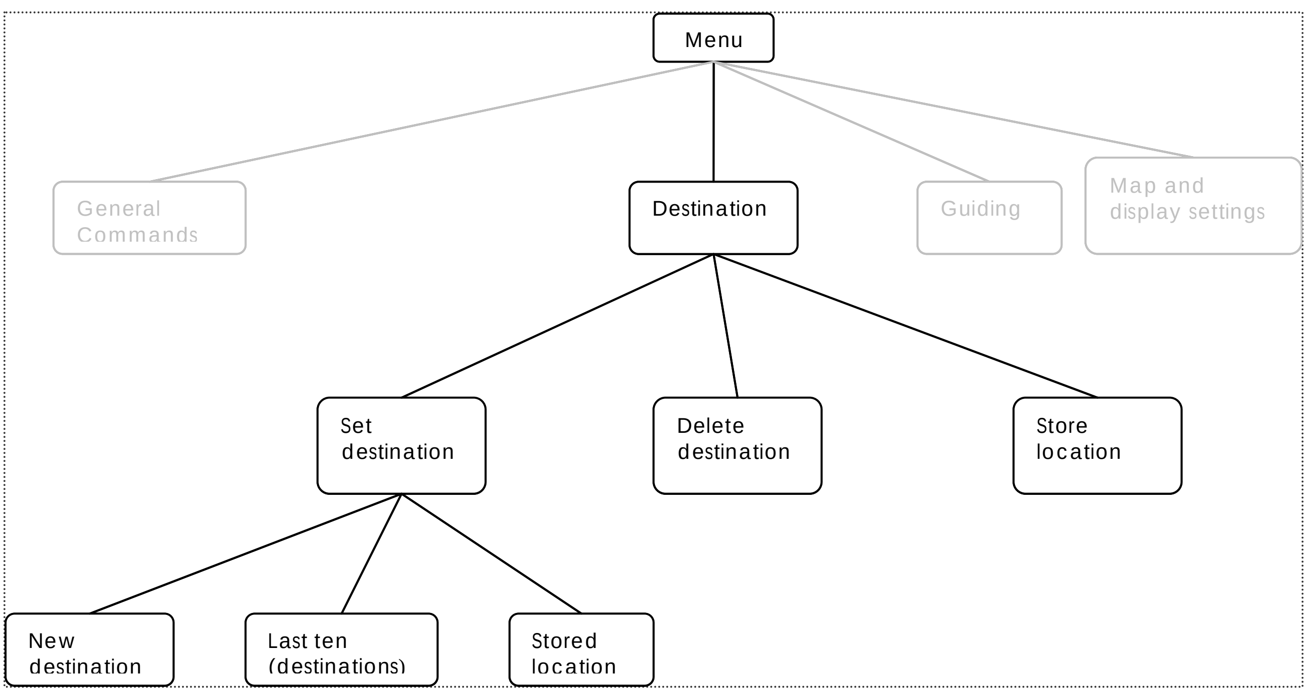


Figure 24: Proposed menu structure, submenu *Destination*.

Destination. In the current help menu (*WCIS*) Guiding and Destination is one category. It is one of the most frequently used categories in the help menu. Since this is a category containing many commands (14) it was divided into two smaller categories; *Destination* and *Guiding*. The category *Destination* is found before *Guiding* because the user has to set a destination before the guiding commands can be used. Functions that are related to setting a destination, deleting a destination and storing a location are placed in this menu. This menu also has a new command; *new destination.* This command is described in the chapter above about implementing multimodality in the system (see chapter 6.1).

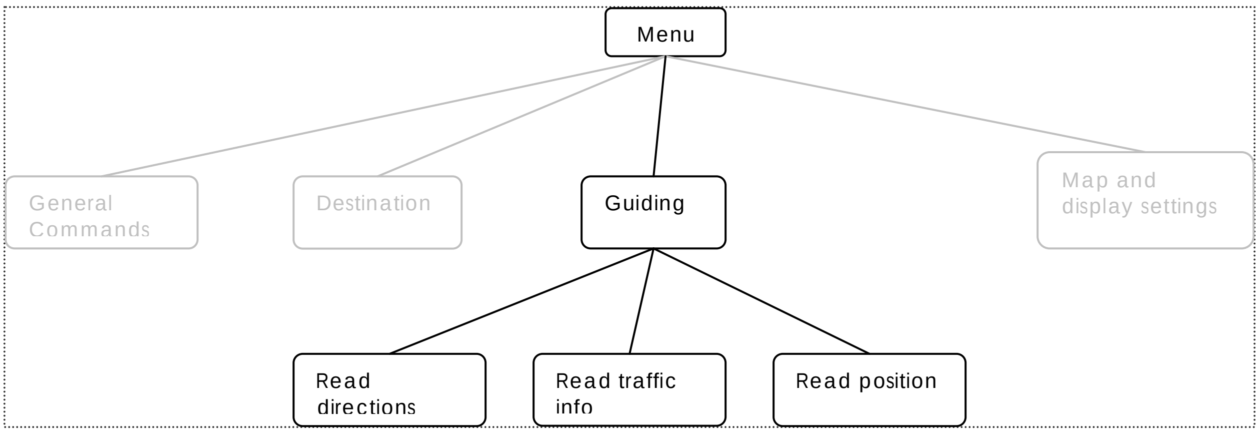


Figure 25: Proposed menu structure, submenu *Guiding*.

Guiding. This part of the menu is the other half of the Guiding and Destination category. The focus of the commands in this category is to aid the user once a destination has been set. The command *read directions* reads out the next guiding directions, *read position* reads out the current position of the vehicle (what road) and both these commands exist in the current prototype. *Read traffic info* is a command that existed in the early version of the prototype but was removed as it caused problems for the users. In user study 2 (see chapter 3.4.2), the respondents seemed to have difficulties in finding the command. Placing the command in this recommended menu structure might make it easier to find but the command should be removed if there is no use for a command like this and if this information is easier to access manually by using the graphical quick menu on the navigation display (this information is available in the current prototype manually).

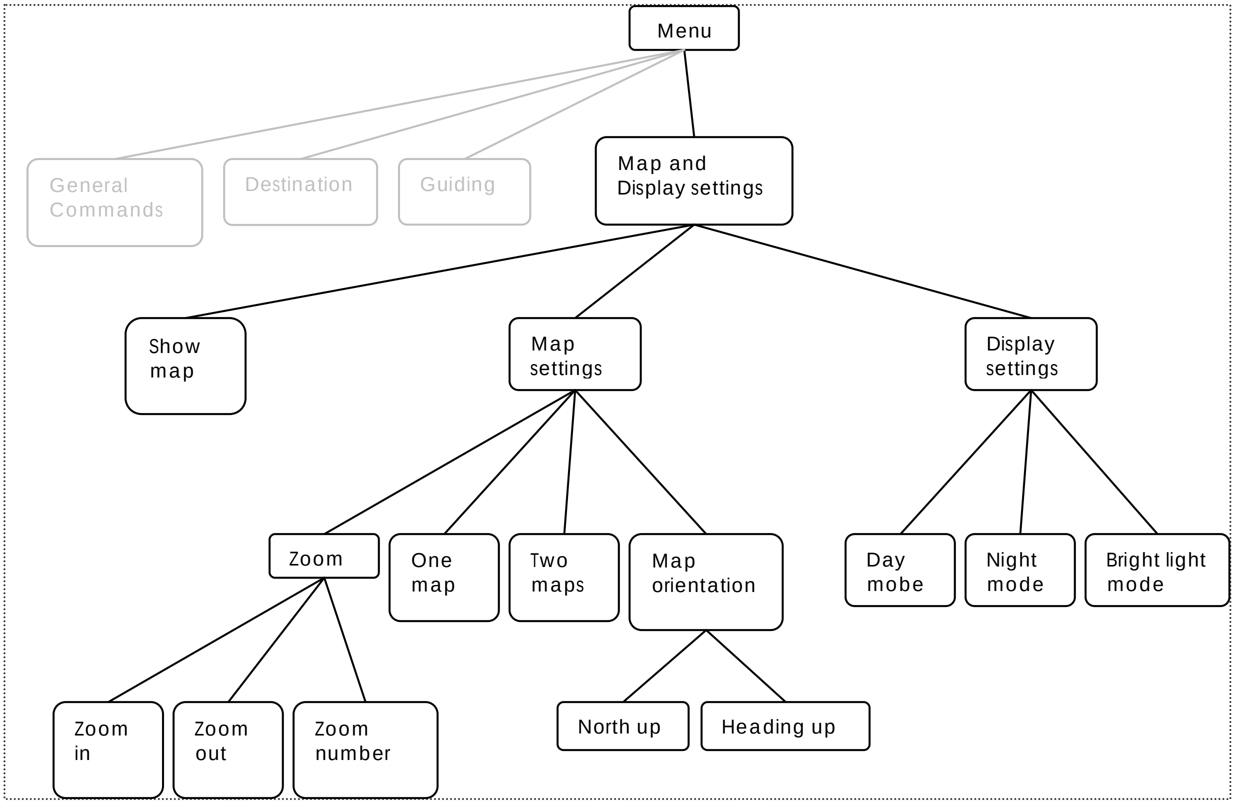


Figure 26: Proposed menu structure, submenu *Map and Display settings*.

Map and Display settings. The current help menu (*WCIS*) has a menu called Map settings. Some commands in the current help menu were setting-commands and have in our recommendation been placed in a new extended settings category called: *Map and Display settings*.

The command s*how map* will show the map on the display; this can be used to clear the display from any graphical menus or hit lists. A new command simply called z*oom* will trigger a dialogue where the system will ask the user if the map should be zoomed in or out, and if the display is showing two maps the user will be asked which map to zoom in/out on (the same way that the dialogue works today). It is also possible to say the commands directly.

The display settings category is new but the commands in it have the same functionality as *bright*, *dark* and *contrast colours*, only the names have been changed. These commands had in the early prototype names similar to the ones recommended but were changed for reasons that are unknown to us. The dark colours (*night mode*) are preferably used for night conditions, the bright colours (*day mode*) for daylight conditions and contrast colours (*bright light mode*) for bright light conditions. We thought it more intuitive to relate to current light conditions rather than trying to remember which colours that are better suited to what light condition.

6.7.2 Redesigning the feedback

Avoid repetitive error messages. Error messages should be redesigned so that they better inform the user about a particular problem and how to solve it. To avoid annoying the users it would be desirable to avoid repetitive error messages. Today when the system can’t recognize a command spoken by the user, one of the error messages “Number or command not recognized” or “Command not recognized” is played. If the user repeatedly fails to pronounce the command he will have to listen to the same message every time. Instead the system could try to help the user e.g. after two failures. By using alternative guesses the system could ask the user “Did you say X or Y?”. The user then answers by saying X or Y. If there is still a problem the system can automatically enter the help function.

Be consistent when giving the user feedback. There is no auditory feedback when a second destination or waypoint has been added to the route, but when a first destination is added the prompt “Please proceed to the indicated route” is played. This makes the users unsure of whether the second destination or waypoint has been added or not. To solve this problem a confirming message could be played every time an additional destination is added. The message could be “Another waypoint has been added to the route”.

Provide the right kind of feedback. The feedback that is given when changing the volume is sometimes difficult to interpret. Feedback is given by playing a tone that varies in volume every time a change has been made (if lowering the volume the tone is played in a softer volume and vice versa). The difficulty lie in interpreting the change in volume (of the played tone) as there is no previous volume to compare with. Sounds are transient, that is not persistent, and therefore it is difficult to decide if a sound is softer or louder than another if not played directly after each other. To solve this problem a tone that grows quieter can indicate that the volume is lowered. A tone that grows louder then signifies that the volume is turned up.

There is no feedback on which volume (navigation or voice control) that has been changed.

Visual/auditory feedback. The feedback when altering between one and two maps is only visual. Feedback is given by the actual change between one or two maps. This is in general a good way of showing that the action has been executed. However it requires the action to be performed immediately which is not the case in the voice controlled navigation system. The lack in speed makes the user repeat the command. Another difficulty is that when the system is about to give a guiding instruction the left half of the display is covered by a route guidance image, showing where to make a turn. When the route guidance image is shown it is not possible to see the change between one and two maps. It can be discussed whether voice feedback should be provided when changing between maps. If in a production system there is still going to be a delay in changing the graphics a prompt saying e.g. “Changing into two maps” could be played. The user would then be reassured that the command has been recognized by the system. This would prevent the user to say the command once again. The best solution would be if the graphics was updated quickly enough to prevent the user from repeating the command.

Redesign the centre stack feedback. People tend to verify more when using a SUI than a GUI. The system should make it easy for the users to verify and check the system’s status of what is happening now and what has happened recently (Weinschenk, 2000). If feedback is given both auditory and visually this reduces the amount of verifications the user will ask for. The visual feedback continues to display thus the user don’t have to remember what was said. Visual feedback should be a complement to the auditory feedback where this is suitable. To enhance the auditory feedback when the PTT-button is pushed, textual feedback is shown on the centre stack display which indicates the voice system’s status. Since half of the test group failed to spot this information, it should be considered to move it elsewhere. The reason for the current position for the feedback on the voice control status is that in the future there is a thought to voice control other systems in addition to the navigation system. Still we suggest moving this feedback closer to the driver’s field of sight to avoid making the driver look down when driving. As suggested by some of the test participants it could be a better idea to use colours instead of textual feedback to inform of the system’s status. LED lamps could be used. When the light is out this can indicate that the PTT-button must be pushed. The colour green could indicate that the system is listening, the colour red that it is processing and thus not taking any further input from the user. There could also be a small icon on the navigation display. The icon of an ear could show that the system is listening. An animated icon of an hour-glass or of a watch could indicate that the system is processing the input.

A better TTS. Feedback given by the current TTS is difficult for the users to perceive. The TTS should preferably be in Swedish for Swedish customers. Today an English TTS is used to read street names in Swedish which makes it sound ridiculous and the output hard to interpret.

The current solution for setting a stored location with voice is today not acceptable. To retrieve a stored location with voice the user has to pronounce the location name (often a street name) as the TTS does. This forces the user to pronounce a Swedish name in English and to pronounce it like a machine. As the user tests showed the output from the TTS was not found satisfying and it is unacceptable to force the users to imitate the TTS output. The user must be able to pronounce the location names in the proper language and this requires a Swedish TTS when using voice control in Swedish.

6.7.3 Redesigning the dialogue structure

Make it easier for the user to reply to the system. Questions about what the user wants to do are in most parts of the dialogue followed by information on what command to use. E.g. “If you want to set a stored location, say stored location”. But when the user is going to set a destination from the list with the ten last destinations he gets the question: "Do you want to set destination to X?" The user is not told how to reply to the question which makes the user hesitant and the user then waits for instructions. This type of question is mostly answered by a yes or no. If the question were spoken with more intonation users would perhaps be more likely to answer it quickly and not hesitate. Another easy way to solve the problem is by applying consistency so that all questions to the user are followed by instructions on how to make a reply; "Do you want to set destination to X? Say yes or no".

Many users overlook the possibility to delete all set destinations in one go as the command *clear itinerary* is presented after "To delete X, say delete destination". Users do not expect another command to follow so they delete all destinations one by one which takes more time. The users have shown to be restless in the interaction with the system. The fact that they have experienced long lists of commands could maybe stress them to choose the first likely command without listening to the following commands as they are afraid to forget the first command. A solution could be to let the system say: "You can delete the destinations one by one or all at once. To delete one destination *say delete destination*, to delete all destinations say clear itinerary" This way of informing the users is similar to how the Jaguar AVP system is designed.

When the user wants to change the volume and the system has recognized which volume to change but not whether to turn it up or down the system prompts: "Please specify; after navigation volume say softer or louder". Even thought the appropriate volume has been identified it is not sufficient to say softer/louder but the entire command is required. A better solution would be if the user only had to say softer or louder.

If the user wants to set a return trip but the destination has been erased the system answers: "No starting point has been set, please set a new destination" The user then believes he can use the command *new destination* to set a new destination which is not possible.

Consistency in how to reply. When entering the function *store location* there are the following suggestions "To store your current position say one. To store your current destination say two". The usual way to answer this type of questions is to specify your choice, in this case by saying *current position* or *current destination.* It is desirable to answer questions with similar structure in a similar way, and thus reserve numbers as commands when choosing from a list with many items.

Speed up reading of lists. It would be desirable to speed up the reading of the street names in the list with the last ten destinations. Many users were impatient when they listened to the list.

6.7.4 Making the system traffic aware

To have a system that is aware of the driver's traffic situation. Volvo has a system called DWE (Driver Workload Evaluator) that calculates the driver's workload from dynamic car information. This system was developed during the EU funded project called Comunicar (Comunicar, 2004). The parameters used for the DWE comes from the five dynamic signal information sources:

- The accelerator pedal position
- The brake pedal position
- The vehicle speed
- The steering wheel angle
- The turn indicator

Volvo also has a similar system in production called Intelligent Driver Information System (IDIS). IDIS consists of two systems; one that calculates workload and one that manages incoming information. IDIS can delay an incoming phone call if the system finds the driver to be experiencing a heavy workload.

It was noticeable during the test that the participants were reluctant to answer any questions or solve any tassk when turning the steering wheel. It seems that situations like these consume most of the driver's concentration and it could be wise to minimize information flow to the user. To minimize the mental workload of the driver a system like IDIS or DWE could be used to halt dialogue prompts from being played until the driver's traffic situation gets easier.

6.7.5 Creating an Introduction CD

When a costumer orders a Volvo car at a car dealer it usually takes some time before it can be delivered. When buying the car the costumer gets an introduction CD with information on the car. One test participant suggested that this CD could contain an introduction to the voice control for the navigation system. Costumers are probably curious about their new car and want to know more about it. This would be a good opportunity to show the costumer how the voice control works and what functions there are. Once the car is delivered the costumer probably tries out the different functions in the car and has not enough patience or interest in reading a handbook or watching a demo.

6.8 ELEMENTS TO KEEP

Even if some users had difficulties to use the PTT-button we recommend keeping the interaction this way as it results in the user being in charge. Another reason for keeping the PTT-button is that most customers of the XC90 are women with children and therefore a noisy environment in the car can be expected. With a PTT-button the driver can choose when the system should listen for commands and e.g. avoid that screaming children confuse the system. Those who had problems with timing when talking to the system improved continuously during the test, it is likely that users quickly learn how to interact with the system by using it. The position of the PTT-button on the steering wheel makes it easy to handle with the thumb the only disadvantage is that when the steering wheel is turned so is the PTT-Button.

Despite the noisy environment the recognition level is over all satisfying but the system should be further trained to recognize the commands where the level of recognition was poor.

Another positive aspect of the system is that it is not chatty except when listening to the long menus in the help function.

Setting commands such as *zoom in* and *one map* are good examples on voice commands that can save time and attention compared to doing the same with manual interaction.

7 DISCUSSION

7.1 ADVANTAGES AND DISADVANTAGES OF THE CO-DISCOVERY TECHNIQUE

The expected advantages of having a co-discovery group was the possibility of gathering more comments from the participants; co-discovery test yields more information about what the users are thinking and what strategies they are using to solve their problems (see chapter 4.5.2). This turned out to be true. The participants of this group gave each other tips and hints on how to solve a task and discussed freely even though some of the pairs of participants had, prior to the test, never met before.

Participant A: *“…säg nått med distance…”*
Participant B: *“… man kan inte bara chansa…”*

In this quote participant B is asked to find out how far they are from the destination and participant B is trying to remember what command to use. Participant A tells him to try to say something with *distance*. Participant B replies that “*you can’t just guess*”. This is a comment that could indicate that participant B is aware that the system is limited in what commands it can recognize. The quote can also indicate that participant A is aware that the system understands stand alone commands (they don’t need to go through the help menu to use the command) and that the command names are related to their functions. They eventually solved the task with some difficulties, this mainly due to the long submenus in the *WCIS* help menu.

The role of the test leader became less central in answering questions asked by the participants, the participants seem to try and solve it on their own, which is what was hoped for. The comments gathered with the co-discovery group tests have been very helpful. They have very clearly pointed out subjective thoughts about the voice control.

On the other hand it was hard to plot problem areas out of the observation notes since the co-discovery group completed most of the tasks. The difference lies in that the participants first discussed how to solve the task and then tried, resulting in a quick completion. Even though the tasks were completed faster the participants spent more time discussing. At one point a participant started taking own initiative on what to do next, without awaiting instructions from the test leader.

The participants who tested individually (the singles group) usually solved the tasks by trial and error. This affected the observation notes. The observation notes gave the illusion of that the participant had greater problems solving the tasks than what they might have experienced.

Participant C: *“Finns det något sätt att hoppa i menyerna?”*

This participant is trying to find a command in the help menu (*WCIS*) and asks the test leader if it’s possible to skip certain parts in the menu. The comment reoccurred several times with other participants who found it tedious to listen through the long submenus. It is not obvious how the participant thinks the task should be solved; the comment does not reveal that.

7.2 USABILITY TEST ON TARGET GROUP

A key element in user centred design is to identify the user. A key element in usability testing is to test the product on the identified user. The main users of the voice controlled navigation system in the XC90 are American stay-at-home-mothers. Our goal was to test the voice control on them but, unfortunately, since the system is not yet out on the market this could not be done out of business confidentiality.

The results may have been affected by the fact that all the participants of the test are employees at Volvo and these have presumably a greater technical knowledge than the average American stay-at-home-mother. Most of the participants had used the navigation system before and one of them had even been involved in the development of it. The fact that not even half the group of participants were female could probably also have affected the test results slightly however it’s impossible to speculate on how.

7.3 VOICE CONTROL TODAY

A couple of years ago voice controlled systems were often just a feature. Because the technology was new and voice control was possible it should be applied to almost anything and everything. Take for instance Nokia’s mobile phones that allowed the user to put voice tags to a phone number. The function was far from perfect because the technology did not allow the user to pronounce the tag differently from the recorded tag. Background noise would also affect the recognition and mobile phones are usually used where there is much background noise. This resulted in a negative attitude towards voice control from the unsatisfied users who could not use the function properly.

Due to extended research and improvement of the voice control technology the recognition is far more advanced today. The user does not have to train a voice recognition system to recognize the users pitch and pronunciation. Unfortunately it is common that people have bad experiences of voice control from early products like Nokia's mobile phone or voice controlled phone services with poor recognition. The demand for voice control systems that are accurate and work smoothly is probably greater today due to past negative experience. Voice control should not be implemented in a system as a feature. To avoid annoying the user voice control should only be used where it improves interaction.

Volvo's plan is to not only let the navigation system be controlled by voice but also to control the audio system and the phone by voice control. It can be questioned which functions in a vehicle that should be voice controlled and which functions benefit from being voice controlled.

8 CONCLUSION

This chapter will debate how and if we have been able to find an answer to our research question.

Looking back at the first part of the research question we asked;

1. *How can an improvement plan to the voice interface prototype operating the navigation system in the Volvo XC90 be made so that usability and safety are at focus?*

The user test that was made in this thesis have laid the ground for the improvement plan, all decisions were made with the material we gathered from the test in mind. The results from the test also inspired to new solutions, as did the cognitive walkthrough of the Jaguar AVP.

To enhance driver safety we focused on making the interaction as smooth and intuitive as possible. This was made by observing the driver’s situation while he was interacting with the navigation system manually and with voice. The structure of the help menu (WCIS) seemed to cause problems for the participants; therefore another menu structure was proposed consisting of a main menu and a help menu.

2. *Could multimodality be implemented in this interface for this purpose?*

During the user test preformed on the Volvo Voice Control Prototype many of the participants tried to use the manual controls to interact with the Speech User Interface or/and asked if it was possible to scroll through the voice menu (*WCIS*) with the manual controls. Most of the participants also seemed positive during the interview towards using multimodality to solve tasks.

The purpose of the voice control system should be to improve the interaction between the user (the driver) and the graphical user interface of the navigation system. That means that only tasks or parts of a task in the navigation system that benefit from being performed with voice should be preformed with voice. It seemed that some tasks were better solved manually, others with voice and some with both modalities.

As a result of this it can be argued that multimodality could be used as a way to improve the interaction.

9 FUTURE WORK

There are four possible tracks for Volvo to follow in the voice project; this thesis work was conducted to help the company choose between the following.

1. Continue the development of the Volvo Voice Control.
2. Use the Jaguar AVP system.
3. Use the Jaguar AVP system and adapt it to Volvo’s navigation system.
4. Abandon the voice project.

We recommend Volvo to follow the first track and to continue to develop the voice control to obtain a product that will meet up with usability requirements and enhance traffic safety. This recommendation is based on conclusions drawn in this thesis and on the aspiration to see a voice control that match the navigation system as much as possible. Already, users have shown that they can interact with the system without great difficulties; they have also manifested a positive attitude towards the system.

We are strongly against following track number two due to the fact that the navigation system and the voice control system would be totally separate. If the voice control is not adapted to the navigation system this would without any doubt limit the systems usability.

As we haven’t been able to test the Jaguar AVP in a car environment we can’t be sure of how the system would perform on the road but there is some doubt about how well the speech recognition would work with the current solution. Therefore track number three is not recommended. Another reason for not choosing that track is that there probably will be great difficulties in adapting the Jaguar AVP to the navigation system as a lot of measures have to be taken. The Jaguar AVP is designed to be more like a separate system for navigation and not as strongly linked to a graphical interface. The Volvo Voice Control, on the other hand, is designed to be an input device to the navigation system and not a system standing alone. To share the resources that a task requires between two areas of the brain as the two Volvo systems do by using both speech and graphics, could perhaps limit the driver’s mental workload. This indicates that it would be preferable to continue the work with the voice control that is already designed for the Volvo’s navigation system.

Abandoning the voice project would imply that a choice that is available in other cars will not be offered in Volvo cars. We believe that voice control can enhance traffic safety if used for making a task easier or less demanding for the mental workload and not only as a funny feature. Using speech technology for the right purposes can result in changing the bad standing of the technology amongst people in their everyday life.

9.1 HOW TO CONTINUE THE DEVELOPMENT OF THE VOLVO VOICE CONTROL SYSTEM

Until today no comparable user study of the Volvo Voice Control Prototype and the Jaguar AVP has been possible to conduct. When the Jaguar system is implemented in a car it would be interesting to make a comparing user study to know more about how the two different interfaces effect the driving. Besides measuring the driver distraction it would be interesting to look into how well the systems recognize and interpret speech input and to see what differences there are in the users' attitudes.

If Volvo decides to allow multimodal interaction with the navigation system, usability tests are recommended to see how well a multimodal system answers to usability requirements.

The results from the user tests showing that users make attempts to interact with the navigation system by using both speech and manual interaction can be input to other studies of multimodal interaction.

The test results also show an interesting aspect on the users need for structure. The participants seemed to have difficulties in getting a clear mental representation of the SUI structure as they entered the help menu (*WCIS*) repeatedly during testing. The participants seemed to think that it was due to their unfamiliarity with the system, that this would improve if they got to test the system more. To test if the learnability of the system is satisfying longer tests should be conducted with the same participants.

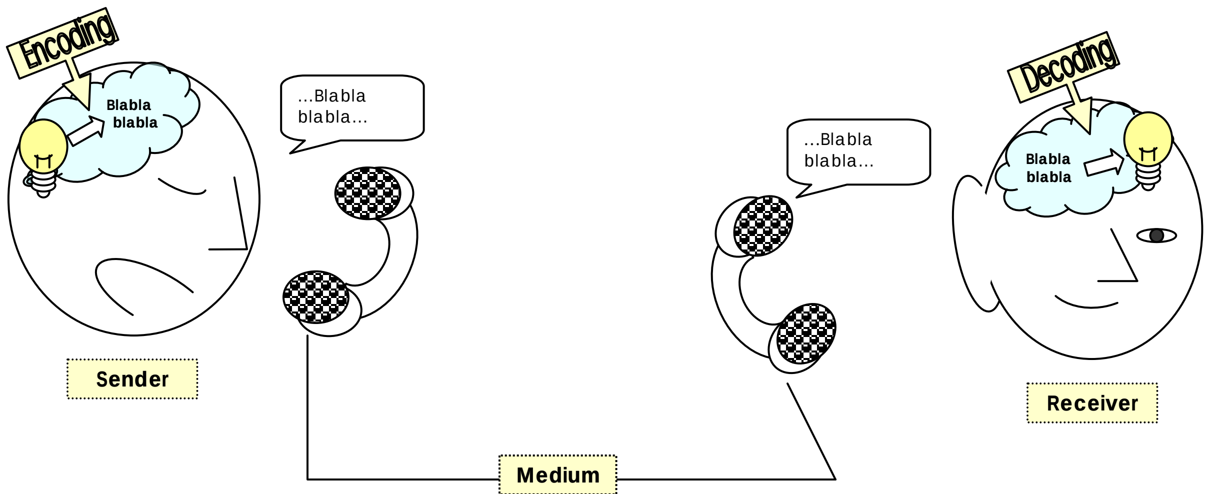
LEXICON

Bottom up and Top down method

Bottom up method
To focus on the subsets of a set, to go from the inside to the outside (Hansen, 1997).

Top down method
To focus on the set followed by the subsets, to go from the outside to the inside (Hansen, 1997).

Communication



According to Herlitz (1989) communication is a process that requires the following:

Sender – is the one who wishes to communicate an idea or a thought. The sender can be a single person or a group of individuals.

Encoding – to create a representation of the idea outside the senders mind a code is used. The code can e.g. be spoken language or a Morse code.

Message – is the encoded idea to be sent. The message is a collection of verbal and nonverbal symbols and holds intended and unintended information from the sender.

Medium – a message is communicated through a medium such as a text, a telephone, spoken language or a computer.

Receiver – the one who receives the message. Also the receiver can be one or multiple individuals.

Decoding – once the message is received by the receiver it is decoded in the mind to give sense to the information in the message.

Context aware
The user gets different messages from the help function depending on where the user is in the software (Schneiderman, 1998).

Dichotomous
A division into two opposed groups or parts.

GUI
A Graphical User Interface is the visual surface of a system that the users experience and interact with.

Haptic device
A manual device that produces sense information e.g. resistance or movements.

Human paced
In a human paced system it is the user who decides on the pace and initiates the interaction.

Mental Workload
Mental Workload is the demands placed on the operator (the one solving the problem) when solving a tasks.

Modality Combinations

Different types of modality can be used together or separately in a system and affect the interaction. Listed below are some types of cooperations between modalities (this chapter on modality combination is gathered from Bunt,1998):

Transfer

> *“When several modalities cooperate by transfer, this means that a chunk of information produced by one modality is used by another modality.” (J.C. Martin, R. Veldman, and D. Béroule, 1998)*

This cooperation of modalities can be used to speed up interaction. As with the case of MAAR (Cheyer and Julia, 1995), where a part of an utterance that has been misrecognized can be edited with the keyboard so that the user does not have to type or utter the entire sentence again.

Equivalence

> *“When several modalities cooperate by equivalence, this means that a chunk of information may be processed as an alternative, by either of them” (J.C. Martin, R. Veldman, and D. Béroule, 1998).*

Using this cooperation of modalities can allow users to choose a modality of preference. In the case of TAPAGE (Faure and Julia, 1994) using a graphical editor the user can choose to specify a command by using either speech or through the selection of a button with a pen. So through equivalence the user is able to select a command with the pen when the speech recognizer is not working accurately because of noise, and thus improving the recognition of the commands.

Specialization

> *“When modalities cooperate by specialization, this means that a specific kind of information is always processed by the same modality” (J.C. Martin, R. Veldman, and D. Béroule, 1998).*

This allowing the information to be presented in the modality that best represents it.

Redundancy

> *"If modalities cooperate by redundancy, this means that the same information is processed by these modalities." (J.C. Martin, R. Veldman, and D. Béroule, 1998)*

Even though redundancy is synonymous with unnecessary, this cooperation of modalities can actually shorten the numbers of steps in an interaction. In a system called COMIT (J.C. Martin, R. Veldman, and D. Béroule, 1998) the user can either quit by saying the command "*quit*" or by typing the command *"quit"*. After that the user will be asked for confirmation. But if the user chooses to both type and utter the command "*quit*" the system will not ask for confirmation, it will simply shut down, hence reducing the numbers of steps in the interaction.

Complementarity

> "*…when modalities cooperate by complementarity, different chunks of information are processed by each modality and have to be merge" (J.C. Martin, R. Veldman, and D. Béroule, 1998).*

This cooperation of modalities can allow the user to speed up the interaction since the interaction can be more intuitive. Take, for instance, multimodal maps where pen and voice cooperate: the user may ask the system *"what is the distance from here to this hotel?"* while simultaneously indicating the specified locations by pointing or circling. This is probably how the user would do if the user was asking another person for directions and pointing on a paper map.

Removable hints

A technique that keeps track on how many times a person has successfully navigated a prompt. Initially the prompt contains hints that suggest things the person can say in response to the prompt. These hints are later on removed when the user knows how to make a reply. This is a technique that provides a novice user with more information than an expert user.

Resource capacity

The performance of every task requires a certain amount of recourses, and the mind is limited to the amount of recourses it can offer at any given time. The mental *capacity* is the maximum or upper limit of processing capability, and *resources* represent the mental effort supplied to improve processing efficiency. There is a relation between the amount of resource the operator can provide and the amount of resources needed to perform the task, which is visualized below (Figure 27) (Wickens, 1992).

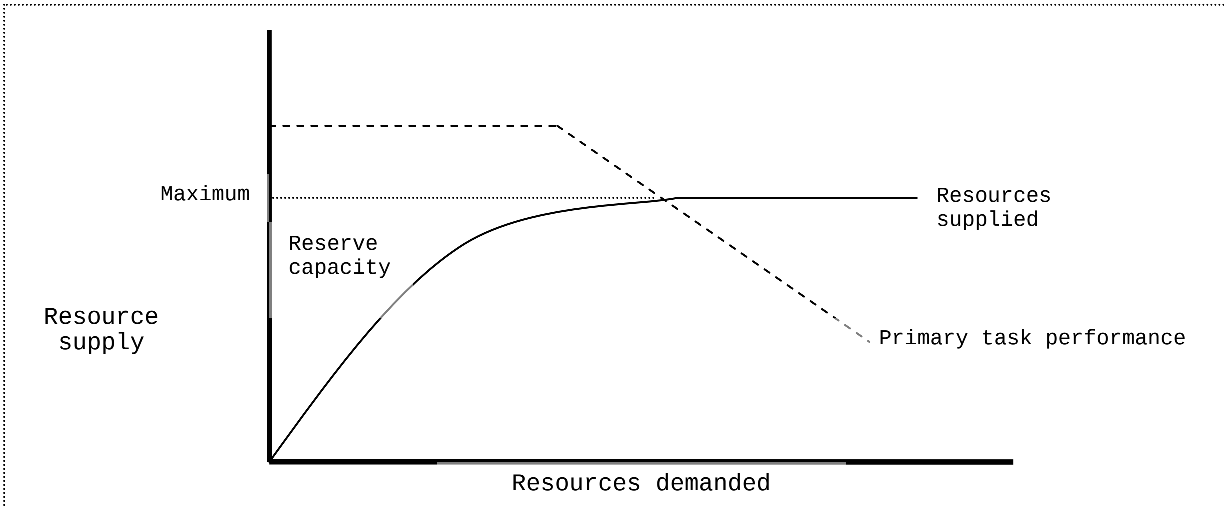


Figure 27 : Shows Wickens model on the relation between the amount of resources the operator can provide and the amount of resources needed to perform a task (Wickens, 1992).

The bottom horizontal line represents the resources demanded by a task and the vertical line represents the resources supplied by the operator (or needed for adequate performance). If adequate performance of a task demands more resources from the operator than available, performance will break down.

The *primary task performance* is a measure used when evaluating a system or an operator. It can be a measure of data-entry speed or driving deviations from the center of the lane (Wickens, 1992).

SUI
A Speech User Interface is the auditory surface of a system that the users experience and interact with.

Voice Control system
A system that makes voice interaction possible.

APPENDIX A

Guidelines for Speech User Interfaces (Weinschenk et al. 2000)

Errors

- Use specific error messages
- Limit background noise
- Allow the user to turn off the input device
- Provide an undo capacity
- Use an auditory icon (to indicate that an error has been made or before an error message is shown or spoken)
- Use multimodal cues for errors if appropriate (If the user interface is multimodal use both modalities to signal an error.)
- Consider offering replay
- Don’t assume people hear everything (put important information first or last)

Feedback

- Supply alternative guesses (of what the user has said)
- Acknowledge the user’s speech (respond with appropriate feedback)
- Show when it is the user’s turn to talk
- Allow for verification (Make it easy for the user to verify and check the status of what is happening and what has happened recently.)
- Use visual feedback
- Use non-speech audio
- Use in-progress messages

Confirmation

- Use confirmations appropriately (Since there are many errors associated with the computer understanding human speech, you may have to use confirmation questions to assure that the computer has heard the right phrase.)
- Ask for clarifying information (if the computer has troubles to understand what the user wants to do)
- Use confirmations for destructive actions
- Be specific (about the information needed by the system)

User expectations

- Identify the computer (let people know that they talk to a computer)
- Build in training time (generally people are not used to speech interfaces)
- Introduce the voice (by providing an introduction or training message)

Keypads and motor actions

- Avoid key combinations (Don’t require the user to press a certain key before or after each command.)
- Use the appropriate word for key-presses (use the word *enter* when the user shall press multiple buttons and *press* for one button)
- Assign command labels consistently (Assign the same key or word to the same function across the system.)

Social environmental issues

- Decide on flexibility (different flexibility for different users and tasks)
- Consider stress
- Consider social interaction (reflect on how the system can affect the interaction as a whole)
- Match the user's work (don't make the user change an optimal work flow to fit the design)

Command-and-control

- Don't use "I" (to avoid that people use longer phrases, more vague requests and are more polite)
- Use modeling (to convey to the user the terminology and structure they should use talking to the system)
- Be brief and terse (if the system speaks in short, terse sentences, the user also will)

Conversation and prompting

- Choose the appropriate word
- Avoid personal pronouns when asking for a response (minimize the number of words)
- Change voices appropriately (Only change voice if there is a change of mode or language. Use a different voice to present warnings.)
- Decide on alerting tones (If synthesized speech is not exclusively used for warnings, then use an alerting tone to precede warnings.)
- Use small steps (speech interfaces tax more on the users memory than graphic interfaces)
- Avoid long prompts or menus (If there is no visual interface attached to the speech interface, users are able to remember about three or four menu or prompt options at a time.)
- Avoid jargon and technical terms in prompts
- Use progressive prompting (Let the systems instructions become more and more specific if the user does not respond.)
- Use prompts to signal going back
- Make prompts directive and exact
- Use an appropriate time-out period and action (usually 10 seconds)
- Allow non-GUI terms (spoken language is less formal than GUI terms on a display)
- Avoid long pauses (long pauses make people uncomfortable and they will want to fill that space)
- Choose an appropriate speed (users will mimic the speed of the computer)
- Use barge-in techniques
- Use tapering (When messages or information has to be repeated over and over. Start with more detailed prompts and then reduce the number of words used.)
- Be consistent
- Use prosodic features (Prosodic features are pitch, volume, speed and pause.)

Menus

- Use result-action for menus with more than two options (First describe the result then the specific action, otherwise the user might forget what action to take.)
- Use action-result for menus with two options
- Let the user know when the menu is complete
- Order the menu appropriately (Put the most frequently-used options at the top of the hierarchy, both in terms of nestling as well as within a particular menu. If a menu item is temporally unavailable, leave it off the menu.)
- Use results as feedback

- Use an appropriate amount of nesting (If the users does not know what they want until they hear the options, then limit the number of menus to four and use nesting; that is sub-menus.)

Non-speech audio/Auditory icons

- Use standard sounds with their usual meanings
- Use volume to signify meaning (A tone that grows quieter indicates that something is going down and a tone that grows louder signifies that something is going up.)

APPENDIX B

Interview questions: American women living in Sweden.
This document has been compressed to minimize the number of pages in this thesis

Interview aim: To investigate driving habits and situations for our Master thesis
Participants can choose not to answer any question.
Age:
Occupation:
For how long time have you been living in Sweden?
Do you have any children? What age are they?

Driving
How often do you drive in Sweden?

Describe a typical driving situation?
Where do you mostly drive?
Short/long distances?
In town? To supermarkets/shops? To work? To kindergarten/school?
In the country-side/outside town? Where? To sport arenas?
On holiday?
Do you often drive to new destinations?
Do you drive alone or with someone else in the car? Who?

How often did you drive in the US?
Where did you mostly drive?
Short/long distances?
In town? To supermarkets/shops? To work? To kindergarten/school?
In the country-side/outside town? Where? To sport arenas?
On holiday?
Did you often drive to new destinations?
Did you drive alone or with someone else in the car? Who?

What main differences are there between driving in Sweden and in the US?

Interests
What do you usually do in your spare time?
What are your interests/hobbies? Sports/museums/restaurants/visits/cinema.

Car & navigation
What car are you driving today?

Does the car have a navigation system?
YES
What do you use it for?
Would you prefer interacting with it manually or with voice?
Why?
NO
Would you consider a navigation system usable?

What would you like to use it for?
Would you prefer interacting with it manually or with voice?
Why?

Computer experience
How often do you use a computer?
What do you use it for?
Have you tried any systems with voice control/voice recognition?

Closing
Would you like to add something to your answers?
Do you have any questions?

Multiple choice question

How would you describe your attitude towards technology?

1. “I find it interesting and I use it as much as I can”
2. “It must be mastered if one is to remain up-to-date”
3. “It is a bit beyond me”
4. “It scares me”

Answers to the interview questions: American women living in Sweden.

This document has been compressed to minimize the number of pages in this thesis. Some participants did not answer every question.

Average Age: 38
Occupations:

- Technical writer
- Language teacher
- Stay at home mom / pre-school teacher
- Home-maker / Marketing

For how long time have you been living in Sweden?

- 4,5 years
- 20 years
- 1 year
- 1 year

Do you have any children? What age are they?

- No children
- 2 children, older than 15
- 3 children, younger than 15
- 2 children, younger than 15

Driving

How often do you drive in Sweden?

- Every day
- 4 days/week
- Every day
- 2 times/week

Describe a typical driving situation?

Where do you mostly drive?

- Pre-school
- To school, work, store

Short/long distances?

- Mostly short distances

In the country-side/outside town?

- 2 women; mostly in the city
- 2 women; both city and outside the city

On holiday?

- In Sweden
- In northern Europe

Do you often drive to new destinations?

- Mostly to known destinations
- Sometimes to new destinations

Do you drive alone or with someone else in the car?

- One woman drives mostly alone
- The other women drive with their husbands/children or alone

How often did you drive in the US?

- Two women said that they drove every day while living in the US.

Where did you mostly drive?

- Two of the women say that the purpose of driving was work-related

- One says that she drove for the same purposes as in Sweden but more vacation trips to visit the family.
- One says that the purpose of driving was to get to the gym, to stores and to teach at her school (work).

Short/long distances?

- Mostly short distances

In the country-side/outside town?

- Both in the city and in the country-side, more in the country-side in the US than in Sweden.

On holiday?

- From Pennsylvania to New York
- More long distances

Did you often drive to new destinations?

- More often to new destinations in the US than in Sweden

Did you drive alone or with someone else in the car? Who?

- Alone or with colleagues

What main differences are there between driving in Sweden and in the US?
Sweden

- More conscious driving
- Less traffic
- More careful driving
- There is no rush hour here
- More difficult to change lane
- More roundabouts
- More sign telling you where you are heading
- More one-way streets

USA

- More signs to indicate where to get off a highway, there are sometimes 4 signs before an exit where in Sweden there are 2.
- Another kind of information on the signs – the signs indicate a direction such as north/south and east/west where in Sweden the signs indicate a route bound for Oslo or the E4 bound for Stockholm.
- More "head on" traffic and less to look out for such as bicycles or pedestrians
- More highways

Interests

What do you usually do in your spare time?
What are your interests/hobbies? Sports/museums/restaurants/visits/cinema.

- Quilting, sports and cinema
- Workout, friends
- Tennis

Car & navigation

What car are you driving today?

- Opel Safira
- Volvo S60
- Volvo V70
- Volvo V70

Some women commented on the fact that they had larger vehicles in the US than what they currently drive in Sweden. The reason for choosing a smaller car in Sweden is due to the difficulty in driving a larger vehicle on Swedish roads; they are smaller and so are the parking lots. This is one of the perceived advantages of having a larger car:

> *"A big car is a good because the children have a good view, there is lot of space so that it's possible to get to the back seat without getting out of the car."*
>
> She had a Chrysler Caravan.

Does the car have a navigation system?

- None of their cars had a navigation system.

NO

Would you consider a navigation system usable?

- Yes
- Yes
- Fun, but too expensive
- Yes, would love one!

What would you like to use it for?

- Travelling
- To find addresses to friends homes
- To find natural fuel stations

Would you prefer interacting with it manually or with voice? Why?

- Voice, it’s safer and easier
- Voice, while driving because I need to concentrate more while driving in Sweden

Computer experience

How often do you use a computer?

- Unanimously: every day

What are you using it for?

- E-mail
- Internet research
- Travel plans
- Work
- Word-processing
- Website

Have you tried any systems with voice control/voice recognition?

- Three of the women have tried voice control.
- For example for telephone services such as banking and purchase of tickets.
- Voice tags in mobile phone.

Multiple choice question

How would you describe your attitude towards technology?

1. **“I find it interesting and I use it as much as I can”**
2. “It must be mastered if one is to remain up-to-date”
3. “It is a bit beyond me”
4. “It scares me”

All the women chose the first option.

APPENDIX C

User and Task Scenarios

Scenario 1, Kate

34 year old Kate lives in the fashionable neighborhood Green Hills outside Washington DC. Kate is a stay at home mother taking care of the house and her children Meg 5, and David 2 years old. This afternoon Kate has to go into town to get her husbands suite at the drycleaner's. Burt, her husband didn't leave it at their usual drycleaner's due to the fact that he was in a hurry. He had forgotten to pick it up and had called Kate and asked her to pick it up for him. Unfortunately Burt had forgotten the name of the cleaner's but he did remember the name of the street, Kings Cross Street. Kate answered that it was no problem as she could use the car's navigation system to find it.

Kate puts the children in the back seat and fastens their seat belts. She enters the car and starts the engine. After having left the garage, she turns on the navigation system and tells it to look for a drycleaner's on Kings Cross Street. Almost directly the system finds *Jim's Dry Cleaning Service* and asks her if she wants to be guided. Kate answers "yes" and the navigation system calculates the route.

Suddenly the system notifies Kate that she's running short on fuel. She is asked the question if she wants the system to guide her to the nearest gas station. Meg and David are by now getting bored by the car ride and start to quarrel loudly in the back seat. Even so the system recognizes her answers. Kate stops to refuel. Once back on the road to the cleaner's Kate remembers that Kings Cross Street has heavy traffic and isn't sure of its parking spaces. She asks the system to plot out available parking spaces around the cleaner's. The system replies that it has found three available places and has plotted them out on the map. She drives to the nearest one and parks the car. The cleaner's is just about to close when she gets there but fortunately she has time to pick her husbands suite up. She wouldn't have liked seeing him wearing jeans and t-shirt at tonight's dinner party….

Scenario 2, Lars

Lars a 43 year old man from Malmö has his own company and spends most of his time on the road. Because he is always on the look for new customers his route isn't always settled beforehand. Today he's going to visit a potential new customer but also an old one. Once on the road Lars wonders who to visit first and figures he might as well let the navigation system calculate the fastest option. His old customer's destination is already added in the navigation system. Lars just adds the new destination and the system calculates that the new client's destination is on the way to the old client's destination. Lars is a very busy man and sometimes forgets to eat. Therefore he also asks the system to plot out restaurants on his way that serves vegetarian food and to plot them out around lunchtime.

Later that afternoon after lunch and the successful visit to his new customer, Lars enjoys the ride and is in the middle of a dialog with the system trying to find out where he could find the nearest pharmacy when a loud bang is heard. The system recognizes the critical situation and pauses the dialogue. Lars pulls over and gets out discovering that his left front tire is flat. Luckily Lars has a spare tire!

Lars has changed the tire and goes back into his car and drives off when he asks the car to continue the dialogue that was paused. The system reminds him that they were looking for a pharmacy and continues the dialogue.

APPENDIX D

Questionnaire for the test participants in the Volvo XC90 test

Enkät för testdeltagare

Vi ber er att fylla i denna enkät och maila tillbaka den till oss så snart som möjligt (AHARUP@volvocars.com). Informationen behöver vi för att dela in er i testgrupper.

Allt material kommer att avidentifieras så fort vi verifierat att alla deltagare finns representerade. Materialet kommer endast behandlas av testledarna.

Namn:

Ålder:

Kön:

Hur många timmar/vecka kör du bil?

Har du en hemdator? (ja/nej)

> **Hur många timmar/vecka använder du den?**
>
> **Vad använder du den till?**

Har du erfarenhet av något röststyrningssystem (exempelvis SJ:s automatiska biljettbokningssystem eller SL:s automatiska tidtabellsupplysnings)? (ja/nej? Om du svarar ja, ange vilka system?)

Har du erfarenhet av navigationssystem i bil? (ja/nej? Om du svarar ja, ange vilka system?)

Answers to the questionnaire sent to the test participants

	Single	**Co-dicovery**
Average age	40	42
Gender		
Male	7	6
female	6	0
Total	**13**	**6**
Drives hours/week		
< 10 -15	5	1
10 - 15	8	3
>10 - 15	0	2
Computer at home		
Yes	13	6
No	0	0
N/A	0	0
Use comp. hours/week		
< 1 hour	4	0
1 - 2 hours	4	2
3 - 4 hours	4	2
> 4 hours	1	2
Use comp. for…		
Online Bank	8	2
Internet (e-mail and info.search)	12	5
Other	4	2
Experience of VC		
Yes	4	0
No	8	6
N/A	1	0
yes; What system?		
	"SJ", Mobile, "Bilprovning", “Västtrafik"	
Experience of Navigation		
Yes	11	6
No	2	0
N/A	0	0

APPENDIX E

Checklist for the user test in the XC90

This observation protocol has been compressed to minimize the number of pages in this thesis

Check list – Före test

- ☐ **Ta fram testpersonernas enkät**
- ☐ **Kolla att körtillstånd finns**
- ☐ **Kolla kamerabatterier**
- ☐ **Kolla att band och extraband finns med**
- ☐ **Kolla bensin**
  - ☐ **Dra streckkod**
  - ☐ **Fyll i tankbok**
- ☐ **Ta bort resmål i navigatorn**
- ☐ **Ställ in guidningsröst till engelsk man**

Introduktion

- ☐ **OK att filma?**
- ☐ **Testupplägg**
  - ☐ **Kort genomgång**
  - ☐ **Körtur med uppgifter**
  - ☐ **Intervju**
- ☐ **Poängtera! Vi testar systemet inte er!**
- ☐ **Visa navigatorn**
  - ☐ **Visa Back- och Enter-knappar**
  - ☐ **Visa hur man får fram**
    - ☐ **Startmenyn**
    - ☐ **Snabbmenyn**
  - ☐ **Visa Voice Control**
    - ☐ **PTT och startljud**
    - ☐ **WCIS som är en hjälpmeny**
    - ☐ **Cancel**
- ☐ **De får testa själva i ca 5 min.**
- ☐ **Poängtera! Det är körningen som är viktigast, inte uppgifterna. De kan alltid be oss att vänta eller repetera uppgiften.**

Check list – Efter test

- ☐ **Skriv rent observationsschema**
- ☐ **Märk band**
- ☐ **Ladda batterier**
- ☐ **Lämna tillbaka nycklar**

APPENDIX F

Task protocol with tasks for the user test in the XC90
This task protocol has been compressed to minimize the number of pages in this thesis

1:a / 2:a

☐ **1. Välj destinationen Temperaturgatan 2 från listan med de 10 sista resmålen** (Voice: *WCIS*→ Set destination→last 10 destinations, Manuellt: Navigation→Set destination→last 10).

☐ ☐ **2. Ta reda på var du befinner dig. M/V** (voice:WCIS→ Guidance and destination→ read position, Utan voice :se karta)

☐ **3. Välj destinationen Barometergatan 1 från listan med de10 sista resmålen Manuell /Voice**

☐ ☐ **4. Byt från en till två kartor M/V** . **Byt tillbaka**(Voice: one map/two maps, Manuellt: ENTER i kartläge→gå till jordglobsikonen, tryck ENTER, välj Two Maps.)

☐ ☐ **5. Sänk volymen för navigationsguiden** (Voice: Navigation volume softer, Manuellt:Navigation→Settings→System options→)

☐ ☐ **6. Ta reda på hur långt det är kvar till målet. M/V** (Voice:WCIS→ Guidance and destination→ remaining route, Manuellt:Står längst upp på skärmen till höger under guidning)

☐ ☐ **7. Höj volymen för navigationsguiden** (Voice: Navigation volume louder, Manuellt:Navigation→Settings→System options→

☐ **8. Ställ in återresa till Assar Gabrielssons väg** (Voice: WCIS→Guidance and destination→Set destination→Return trip, Manuellt: Navigation→Set destination→last 10)

☐ ☐ **9. Ta bort alla destinationer**

Interview questions asked after the user test in the Volvo XC90

This document has been compressed to minimize the total number of pages in this thesis.

Mental modell SUI/GUI

1. Beskriv din bild av systemet. *Hur skulle du/ni beskriva det? Vad kan man göra med det?*

Modalitet

2. Föredrar du/ni generellt sett att göra uppgifterna manuellt eller med röststyrning? **Varför?**
3. Vilket inmatningssätt tycker du/ni passar bäst för vilka uppgifter?

   Varför passar manuell styrning/röststyrning till just denna uppgift?

   Varför passar inte manuell styrning/röststyrning för denna uppgift?
4. Finns det delar av någon uppgift som du/ni skulle vilja kunna göra både manuellt och röststyrt? *T ex. volymen, att man kan anropa specifik volym (navigation volume/voice control volume) med voice men justera med knappar.*

Koppling mellan SUI/GUI

5. Hur vet du/ni när systemet lyssnar på dig/er eller bearbetar vad du/ni har sagt? *Har de sett listening…och processing.* ***Presenteras info på ett bra ställe?***

SUI

6. Kan du/ni förklara hur du/ni gör för att tala till systemet. *Hur fungerar PTT-knappen?* ***Kan man avbryta systemet när det talar? Hur?***
7. Vad anser du/ni om igenkänningen? *Är du/ni nöjd med hur väl systemet uppfattar vad du/ni säger?*
8. Vad anser du/ni om kommandona? ***Är de svåra? För långa? Svåra ord? Svåra att uttala?***
9. Skulle du vilja att systemet har fler moment där det frågar hur det kan hjälpa dig och du svarar, som i en dialog? **Delvis eller helt?**
10. Vad anser du/ni om Hjälp-menyn? *Hittar du/ni information där du/ni tror att du/ni skall hitta den? Är de lagom långa? Speglar menynamnen innehållet väl?*
11. Skulle det hjälpa om Hjälp-menyn visades i navigationsdisplayen?
12. Skulle du/ni vilja kunna pausa en dialog? ***Hur? Vill du göra det med rösten eller manuellt?***

Sammanfattning

13. Bra/dåligt? Förbättringsförslag

APPENDIX G

Observation protocol used during test in the Volvo XC90

This observation protocol has been compressed (from 3 pages to 1) to minimize the number of pages in this thesis.

Name Driver:
Name Passenger :
Time: Date :

Task	M/V	Subject actions	Help given	Comments
1 Tempg				
2 Var ?				

Task	M/V	Subject actions	Help given	Comments
3 Baro-Meterg				
4 1→2 Maps				
5 Volume softer				

Task	M/V	Subject actions	Help given	Comments
6 Rem. dist-ance				
7 Volume Louder				
8 Return trip (Röra)				

APPENDIX H

Observation protocol for the video analysis

This observation protocol has been compressed to minimize the number of pages in this thesis.

Observationsschema för Videointervju

NAMN:

1. Ett system ☐

Två system ☐

2. Manuellt ☐

Voice ☐

3.

	MANUELLT	VOICE
BRA		
DÅLIGT		

4. Volym

JA	
NEJ	

Andra förslag:

5. Har sett display

JA	
NEJ	

Bra placering

JA	
NEJ	

Övriga förslag:

6. Har förstått PTT:

JA	
NEJ	

Har förstått Barge-In

JA	
NEJ	

7.Nöjd med igenkänningen

JA	
NEJ	

Övriga åsikter:

8. Anser att kommandona är

Svåra ord

JA	
NEJ	

För långa

JA	
NEJ	

Svåra att uttala

JA	
NEJ	

Övriga:

9. Fler inslag av dialog

JA	
NEJ	

Varför?:
Var?:

Varför?:

10. Anser sig hitta info där de tror att den ska vara (WCIS)?:

JA	
NEJ	

Anser att menyerna är lagom långa?:

JA	
NEJ	

Övriga åsikter om menyerna:

11. Skulle vilja se WCIS-menyn grafiskt i Navi-display?:

JA	
NEJ	

12.Vill kunna pausa i dialog?:

JA	
NEJ	

Hur?

knapp	
ord	

13.

	MANUELLT	**VOICE**
BRA		
DÅLIGT		
FÖRBÄTTRINGS-FÖRSLAG		

REFERENCES

Ahrenfelt,B., Berner, R. (1999) *Konflikthanteringsboken*, Bäcklunds Boktryckeri Aktiebolag, Malmö,1999.

Allwood, J. (1986) *Mänsklig kommunikation*. Skriftserien Guling. Ulricehamns Tidnings AB:s Tryckeri 1986.

Baber, C., Noyes, J.M. (1993) *Interactive speech technology: Human Factors Issues in the Application of Speech Input/Output to Computers*. Taylor & Francis Ltd, 1993.

Blomberg, M., Elenius, K. (2000) *Automatisk igenkänning av tal.I* nstitutionen för tal musik och hörsel, KTH.

Brems, D., Rabin, M., Wagget, J. (1995) *Using Natural Language Conversations in the User Interface Design of Automatic Speech Recognition Systems*. In Human Factors, 37(2),pp.265-282.

Bunt, H. (1998) *Issues in Multimodal Human-Computer Communication* Computational Linguistics and Artificial Intelligence Group, Tilburg University, 5000 LE Tilburg, The Netherlands, 1998.

Comunicar (2004) Information is found at:
http://www.comunicar-eu.org

Dumas, J.S., Redish, J.C. (1999) *A practical guide to usability testing.* Intellect Books, Portland, Oregon, 1999.

Fant, G. (2000) *Half a century in phonetics and speech research.* Department of Speech, Music and Hearing, KTH, Stockholm.

Faulkner, X. (2000) *Usability Engineering*. Palgrave, 2000.

Gallistel, C.R. (1990) *The Organization of Learning*. The MIT Press/ Bradford Book, (p.35), 1990.

Gardner-Bonneau, D. (1999) *Human factors and voice interactive systems*. Kluwer Academic Publishers, 1999.

Grane, C., Rydström, A. (2002) *Design of a graphical supported in-vehicle speech user interface*. Luleå tekniska universitet, Luleå 2002.
http://epubl.luth.se/1402-1617/2002/367
Hansen, K.B. (1997) Grundläggande logik. Studentlitteratur, Lund 1997.

Heins R., Franzke, M., Durian, M., Bayya, A. (1997) *Turn-Taking as a Design Principle for Barge-In in Spoken Language Systems.* International journal of speech technology 2, pp 155-164, 1997.

Herlitz, G. (1989) *Kulturgrammatik.* Konsultförlaget AB, Uppsala, 1989.

Karlsson, Nina. (2003) Information taken from an internal report at Volvo.

Levitt, T.S., Lawton, D.T. (1990) *Qualitative navigation for mobile robots.* Artificial Intelligence, 44:305-360, 1990.

Lundh, L-G., Montgomery, H. Waern, Y. (1992) *Kognitiv psykologi.* Studentlitteratur, Lund, 1992.

McKinley, A., Arnott J. (1993) *A study of turn-taking in a computer-supported group task.* Found at:
http://citeseer.ist.psu.edu/cache/papers/cs/8760/http:zSzzSzwww.dcs.ed.ac.ukzSzhomezSzrnpzSzPaperszSzhci-93zSzfinal.pdf/mckinlay93study.pdf/

MDI (2002) Power Point presentation from the course "Human Computer Interaction" at the IT-University, Gothenburg 2002. Found at:
http://www.cs.chalmers.se/idc/ituniv/kurser/02/mdi/index.html

Microsoft (2004) The information is found at:
http://www.microsoft.com/speech/evaluation/techover

Roe, D.B., Wilpon, J.G. (1994) *Voice communications between humans and machines*. National Academy of Science, USA, 1994.

Shneiderman, B. (1998) *Designing the User Interface – Strategies for Effective Human-computer Interaction*. Addison Wesley Longman, Inc., 1998.

Taylor, M. M., Néel, F., Bouwhuis, D.G. (2000) *The Structure of Multimodal Dialogue II*. John Benjamins Publishing Company, Philadelphia/Amsterdam, 2000.

Volvo (2004) The information is found at:
http://www.volvocars.com

Weinschenk, S., Barker, D.T. (2000) *Designing effective speech user interfaces*. Wiley, New York, 2000.

Wickens, C.D. (1992) *Engineering psychology and human performance.* HarperCollins, New York, 1992.